FRECHET ET HALBWACHS

LE CALCUL DES PROBABILITÉS

à la portée de tous

DUNOD EDITEUR

PARIS

Purchased by Heman Burr Leonard
June 24, 1931.

LE

CALCUL DES PROBABILITÉS

A LA PORTÉE DE TOUS

PRIX NET SANS MAJORATION

Broché : **18** francs. — Reliure : **2** francs en sus.

E.

LE

CALCUL DES PROBABILITÉS

A LA PORTÉE DE TOUS

PAR

FRECHET ET **HALBWACHS**

Professeur à la Faculté des Sciences — Professeur à la Faculté des Lettres

de l'Université de Strasbourg.

PARIS

DUNOD

92, RUE BONAPARTE (VI)

1924

PRÉFACE

Les applications de la théorie des probabilités sont devenues de plus en plus nombreuses. Objet de curiosité pour les mathématiciens du XVII[e] et du XVIII[e] siècle, envisagée exclusivement dans ses rapports avec les jeux de hasard, où elle servait à déterminer « la règle des parties », elle paraissait encore à Auguste Comte, dans la première moitié du XIX[e] siècle, le type de ces recherches purement spéculatives et un peu vaines où se complaît le byzantinisme de quelques savants. Aujourd'hui, on est frappé au contraire de la place de plus en plus centrale qu'elle occupe dans toutes les sciences théoriques et appliquées, depuis l'astronomie et la chimie physique jusqu'aux différentes sections de la biologie et de la science sociologique. Non seulement tous les expérimentateurs qui ont besoin de déterminer le degré de précision de leurs observations, et de calculer l'étendue des erreurs qu'ils risquent de commettre, doivent s'en pénétrer, mais

elle est à la base de la statistique. C'est dire que le médecin, le démographe, l'économiste, l'actuaire, l'agent d'assurances, etc., s'ils veulent user intelligemment des méthodes dont on leur enseigne la pratique, et posséder « la théorie de leur art », ne peuvent plus l'ignorer.

Cependant, la plupart des traités où elle est présentée supposent des connaissances de mathématiques transcendantes. On y utilise les procédés et les symboles du calcul différentiel et intégral. Le lecteur non initié aux subtilités de l'Analyse supérieure hésite à s'y engager, ou s'arrête bien vite, découragé. Sans doute on a essayé d'en simplifier l'exposé, et on y a réussi dans une certaine mesure. Le grand traité analytique des *Probabilités* de *Laplace*, — l'ouvrage le plus complet qui ait été consacré à cette branche de la science, et la source de tous les développements ultérieurs, — n'est accessible qu'aux mathématiciens. On pouvait en écarter les développements les plus ardus, et en présenter les idées fondamentales sous une forme moins technique : c'est ce que fit Bertrand, avec un talent très personnel, et non sans esprit. Récemment, des auteurs distingués ont exposé l'essentiel de la théorie en des ouvrages qu'un plus grand nombre de lecteurs pouvaient aborder,

quittes à sauter les passages les plus difficiles, ou à compléter au préalable leur culture scientifique. Mais n'était-il pas possible d'aller plus loin, et de composer sur cette théorie un ouvrage qui, en toutes ses parties et d'un bout à l'autre, demeurât à la portée de ceux dont les connaissances mathématiques ne dépassent point le niveau de l'enseignement élémentaire, par exemple celui de nos lycées, collèges et écoles primaires supérieures ?

C'est à cette préoccupation que répond ce livre. Il tire son origine d'une série de leçons sur les probabilités, faites à l'Université de Strasbourg en 1921 par l'un des auteurs, rédigées, complétées, remaniées et mises au point en collaboration avec l'autre. Tous deux, par des voies différentes, avaient été amenés à s'occuper des applications de la théorie des probabilités. Le premier, mathématicien de profession, jugeait que le savant le plus préoccupé de recherches spéculatives ne doit pas se désintéresser de la pratique, et qu'il est d'ailleurs utile au progrès de la science d'en diffuser (nous ne disons point : d'en vulgariser) les résultats. Une partie de ses cours était consacrée à des questions de mathématiques appliquées. Le second, sociologue et statisticien, sentait bien que la méthode statistique n'est qu'une routine pour

qui n'est point capable d'en saisir l'esprit et le sens scientifique profond. Il s'était occupé de bonne heure de plusieurs problèmes que soulève cette théorie, et n'avait pas cessé d'y réfléchir[1]. Chargés d'enseigner, le premier, les assurances, le second, la statistique, à l'Institut commercial d'Enseignement supérieur de Strasbourg, c'est à cette occasion qu'ils eurent l'idée de composer ce livre, où ils présenteraient les principes de la théorie des probabilités et ses applications essentielles en n'utilisant que les notions d'algèbre les plus simples, et même en ne les introduisant que le plus tard possible (on verra que le premier chapitre de cet ouvrage ne suppose que des connaissances primaires, comme les opérations sur les fractions). Ils ont été eux-mêmes surpris, — et peut-être le lecteur partagera-t-il cet étonnement, — de reconnaître à quel point on peut entrer profondément dans la théorie à l'aide de notions mathématiques élémentaires, et quels problèmes difficiles on parvient à résoudre au moyen des seuls résultats établis dans ce livre.

Qu'on nous permette d'ajouter que, s'il vise surtout à rendre plus accessible la théorie des pro-

1. Halbwachs. *La théorie de l'homme moyen. Essai sur Quetelet et la statistique morale.* Paris, Alcan, 1913.

babilités, à cela ne s'est pas limité notre effort. On s'apercevra, nous l'espérons, que, dans le détail comme dans l'ensemble, nous ne nous en sommes pas tenus à suivre les traces de nos devanciers.

Nous devons dire un mot de l'ordre dans lequel nous avons présenté les différentes parties de cette science. M. Lippmann disait un jour à M. Poincaré, à propos du théorème fondamental de la théorie des erreurs : « Tout le monde y croit, parce que les expérimentateurs y voient un théorème de mathématiques, tandis que les mathématiciens le considèrent comme un fait d'expérience. » En est-il de même du théorème des probabilités auquel Bernoulli a attaché son nom ? On s'étonnera peut-être de ce qu'il ne soit point placé, comme dans la plupart des traités, en tête du nôtre, mais seulement dans le sixième et dernier chapitre. Nous n'avons pas, d'ailleurs, à prendre parti entre certains empiristes de valeur, comme Cournot, Venn, Chrystal, etc., et ceux qui font de la Théorie des probabilités une science purement déductive, tirée tout entière de quelques définitions conçues *a priori*. Mais, cherchant un mode d'exposition bien adapté à notre objet, comme nous nous proposions d'utiliser les théorèmes généraux comme autant de règles de calcul numérique en vue de la prévi-

sion d'événements futurs, nous avons convenu de considérer la probabilité comme une notion purement expérimentale, la notion d'une constante physique dont la mesure approchée s'obtient par l'observation positive de la fréquence d'un fait. Il appartiendra à nos lecteurs de reconnaître s'il nous a été possible de nous maintenir à ce point de vue dans tout le cours du livre, et, nous y maintenant, de ne rien sacrifier d'essentiel dans la théorie.

Nous n'avons point parlé de la théorie des erreurs d'observation. Ce n'est pas, bien au contraire, que nous la considérions comme négligeable. Mais nous n'avons pas voulu allonger ce livre. Si le public réserve à celui-ci un accueil favorable, nous nous proposons de l'exposer dans un nouvel ouvrage, où prendrait place également l'étude d'un certain nombre d'applications à la statistique que nous avons dû laisser de côté, par exemple celle de la *covariation*.

Nous nous sommes contentés de proposer quelques exercices d'application après chacun des chapitres, réservant pour plus tard un recueil de solutions (par les méthodes élémentaires indiquées dans notre ouvrage) de la plupart des questions posées aux examens de l'Institut des actuaires

français, de la Statistique générale de la France, du Contrôle des Assurances.

On trouvera, à la fin, sinon une bibliographie complète, peu à sa place dans un livre de ce genre, du moins quelques indications sur les ouvrages dont nous recommandons l'étude à ceux qui voudraient pousser plus avant.

Terminons en assurant nos lecteurs que nous accueillerons avec reconnaissance leurs critiques et leurs suggestions.

Strasbourg, le 7 mars 1923.

LE

CALCUL DES PROBABILITÉS

A LA PORTÉE DE TOUS

INTRODUCTION

FRÉQUENCE. — LOI DU HASARD. — PROBABILITÉ

Lorsqu'on expose la théorie des jeux de hasard, on prend soin, d'ordinaire, d'indiquer à l'avance, et tout au début, ce qu'on entend par le hasard et la probabilité. Mais ces définitions, comme celles qui se trouvent en tête des divers livres de la géométrie classique, sont au fond des conventions ou des hypothèses. Il en résulte que la théorie des jeux de hasard se présente comme une construction de l'esprit. Sans doute les exemples sont bien empruntés à l'expérience. Mais comme les définitions elles-mêmes ne sont pas tirées des faits, rien ne prouve que la théorie s'ajuste au monde réel, et qu'on puisse, par conséquent, en tirer des applications pratiques.

Nous voudrions nous placer ici à un tout autre point de vue. Nous considérons la théorie des probabilités

comme une science positive, où les mathématiques (du moins la partie élémentaire des mathématiques) interviennent, il est vrai, à chaque instant, mais qui doit prendre son point de départ dans un certain nombre de notions de fait, tirées de l'expérience, et qui ne valent que dans la mesure où elles correspondent à des réalités. Du moins il nous paraît y avoir avantage à présenter ainsi ces notions et les développements qui s'y rattachent. C'est pourquoi, toutes les fois qu'il nous sera possible, nous tirerons nos exemples non pas des jeux de hasard, mais des faits naturels enregistrés dans les statistiques.

Tout le monde a le sentiment de ce qu'est le hasard. Quand on dit : « je vous ai rencontré par hasard », ou : « j'ai pris un de ces livres au hasard », on sait, ou du moins on croit savoir ce qu'on entend par là.

Il y a, en d'autres termes, une notion vulgaire et courante de ce qu'est un événement fortuit, notion confuse, bien entendu, et qu'il sera nécessaire de préciser ; mais, si nous partons du donné, il faut bien que nous nous en contentions. Si nous voulions dès maintenant l'analyser et l'expliquer, nous substituerions à ce qui est donné, dans le langage et l'opinion, une formule plus ou moins abstraite, et, inconsciemment, nous introduirions toute une théorie. Donc, nous admettrons que cette notion familière suffit à définir le hasard : c'est sur elle que nous nous appuierons pour former la définition de la probabilité.

La loi du hasard est à la base de la théorie des pro-

babilités. Mais il faut bien comprendre, dès le début et avant d'aller plus avant, en quel sens c'est une loi expérimentale. Nous allons montrer comment on est conduit à la formuler, après qu'on a observé la fréquence de certaines catégories d'événements fortuits (prenant *fortuit*, comme nous l'avons dit, dans l'acception courante de ce terme).

Exemple. — Considérons une statistique où la fréquence d'un même phénomène a été calculée pour des groupes d'individus plus ou moins étendus. Soit la masculinité (c'est-à-dire la proportion des garçons) parmi les mort-nés illégitimes en Alsace-Lorraine. On trouve dans le *Compte Rendu* publié en 1920 par l'*Office de Statistique d'Alsace et de Lorraine* (p. 24) qu'il y a eu en 1913, dans six arrondissements du département du Bas-Rhin, respectivement : 3 ; 6 ; 2 ; 4 ; 14 ; 0 ; 1 garçons parmi les 4 ; 11 ; 3 ; 5 ; 29 ; 2 ; 2 mort-nés illégitimes relevés dans ces arrondissements. Les pourcentages des garçons ont donc été respectivement :

75 ; 55 ; 67 ; 80 ; 48 ; 0 ; 50 ; ou, par ordre de grandeur :

0 ; 48 ; 50 ; 55 ; 67 ; 75 ; 80.

Ces nombres paraissent distribués de façon quelconque, sinon entre les valeurs extrêmes du pourcentage, 0 et 100, du moins dans un intervalle assez étendu.

Mais les groupes que nous venons d'envisager sont peu nombreux [1] (moins de 30 individus pour chacun).

1. Nous appellerons *nombreux* un groupe ou des groupes qui contiennent un grand nombre d'individus ou d'épreuves.

Considérons des groupes plus nombreux, ceux qui correspondent, en 1913 également, à chacun des trois départements : on trouve 30 garçons sur 56 mort-nés illégitimes, 22 garçons sur 36 ; 19 sur 50. Les pourcentages, rangés par ordre de grandeur, sont : 38 ; 54 ; 61. La dispersion est déjà moins grande.

Considérons maintenant le groupe qui correspond à l'ensemble des trois départements, pour chacune des années 1913, 1914, 1915, 1916, 1917, 1918, séparément. On trouve : 71 garçons sur 142 mort-nés illégitimes ; 92 sur 178 ; 72 sur 134 ; 53 sur 102 ; 61 sur 101 ; 65 sur 128, et les pourcentages correspondants rangés par ordre de grandeur sont : 50, 51, 52, 54, 60. Chacun des groupes contient maintenant plus de 100 individus, et on voit que les pourcentages correspondants sont déjà beaucoup plus voisins que dans les très petits groupes considérés d'abord. Ainsi ils tendent à se rapprocher d'une limite déterminée quand on les calcule dans des groupes de plus en plus grands.

Fréquence. — Nous nous sommes servis des mots proportion et pourcentage (ce dernier emprunté à la langue courante : par exemple, quand on dit : tant pour cent). Mais si l'on veut effectuer des calculs précis, ce qui oblige à introduire des décimales au numérateur, mieux vaut parler, par exemple, du rapport annuel du nombre des garçons mort-nés illégitimes dans un groupe au nombre des mort-nés illégitimes (des deux sexes) de ce groupe, ou, encore, du nombre des ouvriers victimes d'un accident, dans un groupe, au

nombre des ouvriers de ce groupe. C'est ce rapport qu'on appelle *la fréquence*. D'où la définition suivante :

Dans un groupe fini d'épreuves, la fréquence d'un événement fortuit est mesurée par le rapport du nombre d'épreuves où cet événement s'est produit [1] au nombre total des épreuves. — Remarquons que cette définition est indépendante de la notion de probabilité.

Si l'on suppose que, dans chacune de ces épreuves, ou bien l'événement se produit, ou bien il ne se produit pas, le nombre des épreuves où cet événement se produit (c'est-à-dire sa répétition) peut être égal tout au plus au nombre des épreuves. Par suite, la fréquence ne peut prendre que des valeurs comprises entre 0 et 1, ces deux nombres inclus, 0 exprimant l'impossibilité et 1 la certitude.

En effet, dans notre exemple, le nombre des garçons dans un groupe de mort-nés ne peut être supérieur à celui des mort-nés de ce groupe.

Donnons d'autres exemples de fréquences :

I. Le grand statisticien belge Quetelet a déterminé par l'expérience la fréquence de l'événement qui consiste à tirer au hasard une boule blanche d'une urne contenant autant de boules blanches que de boules noires (en remettant aussitôt dans l'urne la boule tirée). Il y avait dans l'urne 40 boules blanches et 40 noires de dimensions égales et de forme identique. Il a fait 4.096 ti-

1. Il sera souvent utile d'appeler ce nombre lui-même, *la répétition* de l'événement dans ce groupe d'épreuves.

rages, et a tiré 2.066 fois une blanche et 2.030 fois une noire.

La fréquence de l'événement est donc $\frac{2.066}{4.096}$ c'est-à-dire 0,504. C'est près de $\frac{1}{2}$. — Dans cette expérience, il a considéré les résultats obtenus pour des groupes comprenant sept tirages seulement. Les fréquences ne pouvaient être que 0, $\frac{1}{7}$, $\frac{2}{7}$, $\frac{3}{7}$, $\frac{4}{7}$, $\frac{5}{7}$, $\frac{6}{7}$, 1, suivant qu'au cours des 7 tirages on tirait 0, 1, 2 ..., 7 boules blanches. En fait il a trouvé, pour 585 groupes de 7 tirages, les proportions suivantes :

PROPORTION DES GROUPES DE 7 BOULES :							
2,0	7,4	22,7	33,0	32,4	20,8	8,8	0,9
OÙ LA FRÉQUENCE DES BOULES BLANCHES EST :							
0	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{4}{7}$	$\frac{5}{7}$	$\frac{6}{7}$	1

La fréquence $\frac{1}{2}$ correspond à $\frac{3\frac{1}{2}}{7}$, et on voit que le plus grand nombre des groupes ont donné une fréquence égale à $\frac{3}{7}$ ou $\frac{4}{7}$, 0,43 ou 0,57.

II. Le statisticien allemand Lexis a calculé la fréquence des naissances masculines par rapport au nombre total

des naissances, en considérant séparément 816 groupes de naissances en Prusse la même année.

Les chiffres qu'il donne représentent non la fréquence elle-même, mais le rapport de la fréquence m des naissances masculines à la fréquence f des naissances féminines. Il est évident que s'il naît, par exemple, pour un total de 100 naissances, 51 garçons contre 49 filles, on aura $m = \frac{51}{100}$, $f = \frac{49}{100}$, d'où $m + f = \frac{51 + 49}{100} = 1$ (D'une façon générale il est évident que la somme de la fréquence d'un événement et de l'événement contraire est égale à l'unité). Lexis a donc calculé $m = \frac{m}{1 - m}$.

Il a trouvé que le rapport $\frac{m}{1 - m}$ était égal à 1,063 pour l'ensemble total des naissances des 816 groupes, qu'il n'était pas exactement le même dans chacun des groupes, mais qu'il différait très peu, dans la plupart des groupes, du nombre trouvé pour l'ensemble.

NOMBRE des groupes	OÙ LA DIFFÉRENCE entre le rapport $\frac{m}{1 - m}$ et 1,063 était de :	C'EST-A-DIRE OÙ : pour 1.000 naissances féminines, il y a un nombre de naissances masculines qui diffère de 1.063 de :
282	0 à 0,0195	0 à 19 1/2
214	0,0195 à 0,0395	19 1/2 à 39 1/2
135	0,0395 à 0,0595	39 1/2 à 59 1/2
94	0,0595 à 0,0795	59 1/2 à 79 1/2
47	0,0795 à 0,0995	79 1/2 à 99 1/2
44	plus de 0,0995	plus de 99 1/2

Autrement dit, dans le plus grand nombre des groupes le rapport $\frac{m}{1-m}$ variait de 1,02 à 1,10, et par conséquent la fréquence m variait seulement de 0,505 à 0,523.

Ainsi la fréquence d'un événement *fortuit* dans une catégorie d'épreuves déterminée reste sensiblement la même quel que soit ce groupe s'il est très nombreux. Quand le groupe n'est pas très nombreux, les fréquences qu'on rencontre le plus habituellement sont celles qui diffèrent le moins de la fréquence du même événement dans un groupe très nombreux. C'est donc un nombre susceptible d'une mesure expérimentale.

Ces exemples suffisent pour faire comprendre la loi générale suivante :

Loi expérimentale du hasard. — *Pratiquement, lorsque, dans une catégorie d'épreuves déterminée, un événement fortuit se produit un certain nombre de fois, la fréquence de cet événement pour des groupes très nombreux (c'est-à-dire très grands) varie peu. Le nombre d'épreuves où l'événement se produit (la répétition de cet événement) reste en général sensiblement proportionnelle au nombre des épreuves.*

Les variations entre les fréquences observées sont généralement d'autant plus faibles que les groupes sont plus nombreux. — Ce dernier énoncé reste un peu vague. Nous le préciserons plus tard en nous appuyant sur le théorème de Bernoulli [1].

1. L'intérêt de ce théorème ne consiste pas, comme on le croit parfois, en ce qu'il servirait à démontrer la loi du hasard. Celle-

Probabilité. — On peut alors définir *la probabilité d'un événement dans une catégorie d'épreuves* comme un nombre dont on obtient une valeur approchée en calculant la fréquence de cet événement dans un groupe d'épreuves pris au hasard dans la catégorie considérée. La loi du hasard nous enseigne que la précision de cette mesure devient pratiquement illimitée quand le nombre des épreuves de ce groupe croît indéfiniment.

On a ainsi une notion analogue à celle de la mesure d'une grandeur physique quelconque. Si, par exemple, on mesure un angle astronomique successivement avec divers instruments gradués les uns en degrés, les autres en minutes, les autres en secondes, on n'arrive pas toujours aux mêmes résultats, mais ces résultats diffèrent assez peu d'un même nombre qu'on prendra pour mesure de cet angle. Ces résultats différeront d'autant moins que la graduation sera plus précise, de même que les fréquences dans divers groupes différeront d'autant moins que ces groupes comprendront chacun un plus grand nombre d'épreuves.

On pourrait objecter que les catégories d'épreuves étudiées en statistique ne peuvent comporter qu'un nombre fini d'épreuves. Par exemple la probabilité de vivre une année de plus à partir de 45 ans ne peut se mesurer que sur des groupes d'hommes nécessairement au plus

ci est une loi expérimentale, et ne peut se démontrer mathématiquement. Mais, — quand on y joint la formule de Laplace — ce théorème permet de déterminer avec précision l'écart probable entre les fréquences de l'événement dans les différents groupes, et la fréquence de l'événement dans l'ensemble des groupes.

aussi nombreux que la population totale de la terre. Mais la même objection pourrait être faite dans le cas de la mesure d'une grandeur physique quelconque. On n'a jamais fait un nombre infini de telles mesures. Il suffira donc que les fréquences observées se groupent autour d'une même valeur, et d'autant plus que les groupes où on les constate comportent chacun un plus grand nombre d'épreuves.

Il est très important, pour définir la probabilité d'un événement, de bien déterminer la *catégorie d'épreuves* à laquelle s'applique cette définition. Suivant la façon dont sera délimitée cette catégorie, la probabilité sera très différente.

Si je demande la probabilité qu'il y a pour que Pierre meure dans l'année, la question reste indéterminée. Il faut que j'indique sous quel aspect j'envisage Pierre : comme un mammifère, comme un homme, comme un Français moyen, ou comme un forgeron moyen, ou comme un homme de 35 ans, ou comme un forgeron français de 35 ans, etc., c'est-à-dire que je précise la catégorie où Pierre est supposé pris au hasard. Suivant que je le prends dans telle ou telle, la probabilité variera, ou n'aura aucune valeur déterminée.

On dit quelquefois que la probabilité est relative en partie à notre connaissance, en partie à notre ignorance. C'est exprimer assez mal l'idée que nous venons d'indiquer. Car nous pourrions peut-être donner sur Pierre d'autres détails (que nous connaissons) : mais il faut s'arrêter, car si nous multiplions les caractéristiques

de Pierre, la catégorie se réduira à Pierre lui-même, ou à un groupe trop peu nombreux pour qu'il puisse être question de probabilité.

Il n'est donc pas exact que la détermination de la probabilité doive être d'autant plus précise que notre ignorance est plus complète. Quand nous ne tenons pas compte de certaines caractéristiques pour déterminer la catégorie d'épreuves, ce n'est pas nécessairement parce nous ne les connaissons pas.

EXERCICES

I (F)[1]. — Le tableau suivant[2] fournit le relevé en 1907 et en 1912 des nombres (en milliers) des successions françaises réparties suivant le montant de l'actif net :

	1907	1912
francs		
1 à 500	116,3	103,1
501 à 2.000	106,8	93,8
2.001 à 10.000	114,7	101,0
10.001 à 50.000	48,0	45,8
50.001 à 100.000	7,7	7,7
100.001 à 1.000.000	7,6	7,0
Plus d'un million	0,5	0,5
Total	401,6	358,9

1. Dans la suite, nous indiquerons au moyen de la lettre (F) les exercices les plus faciles.

2. Voir à la page 113. XXXVI^e volume, 1919-1920, de l'*Annuaire Statistique* publié par le service de la Statistique générale de la France. Nous désignerons dans là suite, en abrégé ce même volume par « Statist. gé. Fr. »

Calculer les pourcentages annuels des nombres des successions des diverses catégories. Vérifier que ces pourcentages varient peu de 1907 à 1912.

II (F)[1]. — Le tableau suivant[2] fournit, pour chaque millésime, le nombre (en milliers) $\frac{n}{1.000}$ de mineurs travaillant dans les mines et carrières de France (mines de houille exceptées) et le nombre r des mineurs tués.

MILLÉSIME	$\frac{n}{1.000}$	r
1891	128	137
1892	129	160
1893	129	157
1894	130	175
1895	135	158
1896	142	164
1897	144	176
1898	146	183
1899	150	196
1900	150	207

Calculer pour chaque année la fréquence des accidents mortels parmi les mineurs (on se contentera d'évaluer ces fréquences annuelles à un dix-millième près, c'est-à-dire qu'on s'arrêtera à la quatrième décimale).

1. Voir note 1, page précédente.

2. *Statist. gé. Fr.*, page 149.

CHAPITRE I

COMBINAISONS DES PROBABILITÉS

Si l'on se propose de rechercher quelle est la probabilité d'un événement fortuit, on peut appliquer la méthode directe qui consiste à la tirer d'une statistique aussi étendue que possible, c'est-à-dire à compter combien de fois cet événement s'est répété.

Mais il peut arriver que cette statistique soit difficile à établir, et que le dénombrement soit trop long. Il y aura alors avantage à calculer indirectement la probabilité au moyen d'un petit nombre de propositions que nous allons établir.

Théorème de la probabilité totale. — On a dressé des statistiques qui donnent les tailles des conscrits en divers pays. En France, par exemple[1], en 1912, la fréquence des conscrits dont la taille était inférieure à $1^m,54$ était 0,018 (c'est-à-dire que sur 1.000 conscrits 18 avaient une taille inférieure à $1^m,54$). Celle des conscrits dont la taille était comprise entre $1^m,54$ et $1^m,62$

1. *Stat. g. de la Fr.*, page 18.

était 0,262. Ces deux événements s'excluent réciproquement (ceci à condition qu'aucun conscrit n'ait exactement $1^m,54$ ou qu'on ait convenu par exemple de placer ceux-ci dans la deuxième liste, ce qu'on a très certainement fait). Ceci étant, si on considère l'événement qui consiste en ce qu'un conscrit ait une taille inférieure à $1^m,62$, on peut dire que les deux premiers événements (taille inférieure à $1^m,54$, taille comprise entre $1^m,54$ et $1^m,62$) sont deux manières dont peut se produire le dernier (taille inférieure à $1^m,62$) qui s'excluent mutuellement, et qui sont les deux seules manières dont il puisse se produire. Or, sur 1.000 conscrits le nombre de ceux dont la taille était inférieure à $1^m,62$ était évidemment égal à $262 + 18 = 280$. Donc la fréquence du dernier événement est la somme des fréquences des deux premiers

$$0,28 = 0,262 + 0,018.$$

Le raisonnement fait sur l'exemple précédent s'étend sans peine au cas général.

Considérons une certaine catégorie d'épreuves (possibles). J'y prends un groupe déterminé et nombreux de N épreuves (qui ont réellement eu lieu). Soit un événement E, qui peut avoir lieu de deux manières qui s'excluent mutuellement E_1 et E_2. Dans les N épreuves, E_1 se sera produit m_1 fois, E_2, m_2 fois. Combien de fois se sera produit E? $m_1 + m_2$ fois. La fréquence de E_1 ou

$$\frac{m}{N} = \frac{m_1 + m_2}{N} = \frac{m_1}{N} + \frac{m_2}{N}.$$

Si j'appelle F, F_1, F_2 les fréquences respectives de E, E_1, E_2, j'aurai $F = F_1 + F_2$.

Appelons p, p_1, p_2 les probabilités respectives de E, E_1, E_2. Supposons que N devienne très grand : F, F_1, F_2 seront des mesures expérimentales approchées de p, p_1, p_2. A la limite, on aura : $p = p_1 + p_2$.

Le même raisonnement s'applique aux cas où l'événement E se réalise de plus de deux manières, pourvu que ce soit un nombre k fini : $p = p_1 + p_2 \dots + p_k$.

D'où l'énoncé général :

Lorsqu'un événement fortuit peut se produire de plusieurs manières qui s'excluent mutuellement, sa probabilité est égale à la somme des probabilités de chacun de ces modes de production de l'événement (les diverses probabilités envisagées étant supposées relatives à la même catégorie d'épreuves).

La démonstration générale suppose essentiellement que les événements envisagés satisfont bien à la loi du hasard, c'est-à-dire suppose l'existence d'un nombre, appelé probabilité, dont les fréquences dans des groupes très nombreux quelconques sont voisines. L'application du théorème aux cas concrets n'est exacte que dans la mesure où les fréquences dont on se sert pour mesurer les probabilités en sont suffisamment rapprochées, c'est-à-dire ont été calculées pour des groupes assez nombreux. Cette condition est vérifiée ici parce que le nombre des conscrits en France est très grand, et que les groupes pour lesquels on a calculé les fréquences d'une même taille y ont été pris au hasard. Ce ne serait pas le cas si on

prenait le groupe dans un département déterminé, où la taille pourrait être par exemple plus élevée que dans le reste de la France.

Corollaire. — La probabilité d'un événement fortuit est égale au nombre obtenu en retranchant de l'unité la probabilité de l'événement contraire.

Si E est certain, sa fréquence égale 1, et la limite de cette fréquence, ou sa probabilité, égale 1.

Si E est impossible, sa fréquence égale 0, et la limite de cette fréquence, ou sa probabilité, égale 0.

Soit E un événement fortuit, H l'événement contraire. Dans une épreuve quelconque, on trouvera ou E, ou H. Telle est la certitude, dont la probabilité égale 1 ou = probabilité de E + probabilité de H. Appelons-les p et q.

$$p + q = 1 \qquad p \quad 1 - q.$$

Exemple. — Soit, dans une population déterminée, p_x la probabilité, pour une personne d'âge x, d'être vivante au bout d'une année.

Soit q_x la probabilité, pour une personne d'âge x, de mourir avant la fin d'une année.

On aura

$$p_x = 1 - q_x.$$

Cas favorables. — On appelle *cas favorables* ceux où l'événement se produit, *cas défavorables*, ceux où il ne se produit pas. Il y a d'ailleurs, pour un événement, plusieurs manières également de se produire ou ne pas se produire. Par exemple, au jeu de dés, si l'événement consiste à amener un point impair, on peut distin-

guer trois manières d'obtenir un cas défavorable, savoir : 2, 4 ou 6.

Deux cas d'un même événement sont *également possibles* s'ils ont la même probabilité. Par exemple, de la statistique des tailles des conscrits il résulte que la fréquence des tailles de $1^m,64$ et celle des tailles de $1^m,66$ sont toutes deux égales à 0,065. On peut dire qu'il y a autant de chances pour un conscrit d'avoir $1^m,64$ ou d'avoir $1^m,66$ ou encore que ces deux cas sont également possibles.

Avant de tirer les conséquences de ces définitions, il est essentiel d'attirer l'attention sur la distinction *souvent négligée* et pourtant fondamentale entre *cas* favorable et *épreuve* favorable, *cas* défavorable et *épreuve* défavorable. Contrairement à la notion d'épreuve : jet d'un dé, constatation si, au bout d'une année, une personne déterminée est en vie ou non, etc., la notion de cas (favorable ou défavorable) est purement subjective : elle désigne une des modalités sous laquelle, pour la commodité de nos raisonnements, il nous plaît d'envisager l'accomplissement de l'épreuve. Par exemple l'épreuve consiste à tirer une carte, l'événement fortuit à tirer un as de carreau. On peut dire, si l'on veut : il y a deux cas : tirer une carte rouge ou une carte noire, le premier parfois favorable, l'autre défavorable. Mais on peut dire aussi bien, il y a huit cas : tirer un as, un sept, un huit, un neuf, un dix, un valet, une dame, un roi, le premier parfois favorable, les autres toujours défavorable.

Corollaire. — Lorsqu'un événement déterminé E peut se présenter suivant un nombre entier F *de cas favorables également possibles* qui sont les seuls possibles et qui s'excluent mutuellement, la probabilité de E est égale à F fois la probabilité commune à chacun de ces cas.

Corollaire. — Si, pour une certaine catégorie d'épreuves, tous les cas possibles peuvent être répartis en un nombre fini de cas également possibles et s'excluant mutuellement, pour chacun desquels on peut préciser si un certain événement E se produit ou non, la probabilité de E dans cette catégorie d'épreuves est égale au quotient du nombre F des cas favorables par le nombre P des cas également possibles.

Cette proposition est une simple conséquence de la définition statistique de la probabilité et du théorème de la probabilité totale. Car la probabilité de chaque cas est $\frac{1}{P}$, puisque P fois celle-ci égale 1, et alors, d'après le corollaire précédent, la probabilité de E est $\frac{F}{P}$.

En général, on la présente au contraire dès le début comme une définition abstraite de la probabilité, sous la forme : la probabilité d'un événement est le quotient du nombre des cas favorables par le nombre des cas également possibles.

Naturellement il faut alors définir l'égale possibilité sans faire intervenir la probabilité, sans quoi on commettrait un cercle vicieux. On dira, avec Bernoulli et Laplace, que deux cas sont également possibles lorsque,

avant l'événement, on ne voit pas de raison pour que l'un se produise plutôt que l'autre, ou, avec Poincaré, lorsqu'ils sont également vraisemblables.

Dans certaines applications très spéciales du calcul des probabilités (en particulier dans la théorie des jeux) il est en effet facile de répartir les cas en catégories également vraisemblables. Par exemple, on jette en l'air une pièce de monnaie. On dira qu'on ne voit pas de raison *a priori* pour amener plutôt pile que face. Pourquoi ? S'appuie-t-on sur un raisonnement logique ? Mais supposons que nous jetions ainsi des pièces de plus en plus grandes et de plus en plus lourdes. Il arrivera un moment où nous aurons le sentiment très net, une fois que nous lâcherons la pièce, devenue maintenant un disque de métal d'un diamètre et d'une épaisseur considérables, qu'elle va retomber sur un côté plutôt que sur un autre. Or c'est l'expérience qui autorise notre prévision. Mais pourquoi ne nous aurait-elle pas appris également que des pièces de dimensions plus petites, lorsqu'on les jette, tombent à peu près un nombre égal de fois sur une face et sur l'autre ? Puisque le raisonnement ne vaut que dans certaines limites expérimentales, c'est sur l'expérience, expérience vague, trop souvent répétée peut-être pour que nous en gardions un souvenir précis, que le raisonnement lui-même doit reposer. Pour le moment, contentons-nous de constater que, dans tous les cas classiques où on a été amené à considérer deux cas comme également vraisemblables, l'expérience a vérifié cette conclusion. Reste à savoir si

la conclusion ne résultait pas elle-même d'observations inconscientes.

En réalité le principe, implicitement admis en général, que deux cas également vraisemblables *a priori* dans une même catégorie d'épreuves ont la même probabilité statistique, ce principe n'est qu'une hypothèse. Il faudra ou bien mesurer les fréquences des deux cas dans des épreuves nombreuses : si on vérifie qu'elles sont sensiblement égales, on pourra admettre que leurs probabilités sont égales ; ou bien, si on raisonne directement sans expériences préalables, il faudra vérifier par des expériences nombreuses si les probabilités d'événements complexes qu'on en déduit se rapprochent des fréquences réelles de ces événements.

Voici quelques vérifications de ce genre. Si une pièce de monnaie est bien régulière, amener face ou amener pile paraissent deux événements également vraisemblables. Il en est de même du tirage d'une boule blanche ou d'une boule noire dans une urne contenant un nombre égal de boules (identiques) de chaque couleur. De même, amener aux dés chacun des six points 1, 2, 3, 4, 5, 6 paraît également vraisemblable, si les dimensions des trous qui marquent les points dans le dé ne sont pas telles (si les trous ne sont pas si profonds) qu'ils détruisent la symétrie du dé, etc. Voyons si l'expérience vérifie ces prévisions.

Au jeu de pile ou face, Buffon a jeté 4.040 fois une pièce, et a obtenu 2.048 fois face, ce qui donne pour face une fréquence de 0,507, pour pile, de 0,493.

L'écart, quoique petit $\left(\frac{7}{1.000}\right)$ a été jugé exceptionnel par Poisson. Pour les tirages de boules d'une urne, nous avons rappelé (page 6) les expériences de Quetelet qui ont donné comme valeur 0,496 et 0,504, valeurs égales à $\frac{8}{1.000}$ près. Pour le jeu de dés, W. F. R. Weldon a jeté 4.096 fois douze dés, et a obtenu, dans ces 4.096×12 épreuves, une fréquence du point 6 égale à $\frac{1}{6}$ au 10.000e près, ce qui est conforme aux prévisions. Pour un jeu de cartes bien battu, les événements consistant en ce qu'est retournée une carte déterminée paraissent également vraisemblables, et l'expérience a montré que, dans un grand nombre d'épreuves, leurs fréquences sont sensiblement égales, et par conséquent, d'après le corollaire de la page 18, égales à $\frac{1}{32}$ pour un jeu de 32 cartes.

De même aussi, au jeu de la rouge et de la noire, on constate que les événements qui consistent dans l'arrêt d'une petite flèche devant un trait rouge ou un trait noir se répètent à peu près aussi souvent, et qu'en fait leur probabilité est $\frac{1}{2}$. Pour un petit nombre d'expériences il y aura sans doute des écarts assez grands. Mais on peut observer par exemple, suivant M. de Montessus de Ballore, qu'à Monaco on n'a jamais vu sortir trente fois de suite la même couleur.

Il semble que les probabilités *a priori*, si elles sont confirmées par l'expérience, présentent ce caractère que leur valeur, donnée *a priori*, est indépendante du temps,

si les considérations qui l'ont fait deviner en sont elles-mêmes indépendantes. Au contraire, si on considère les probabilités statistiques qui se rapportent à la vie sociale, il est possible que des groupes très nombreux pris dans une même catégorie donnent des fréquences très voisines à la même époque, mais dont la valeur commune varie lentement avec le temps. Il en est ainsi par exemple des tables de mortalité. Mais comme rien ne prouve que les propriétés les plus constantes en apparence de la matière inanimée ne varient point par degrés insensibles, toute probabilité n'a de valeur que dans les limites d'une expérience dont la durée peut être très grande, mais n'est jamais infinie. La différence à cet égard entre les deux espèces de probabilités est toute relative.

Théorème des probabilités composées. — Considérons les tables de mortalité en France, en 1901. Sur 68.837 personnes âgées de 30 ans, 63.082 ont atteint 40 ans, et 56.057 ont atteint 50 ans.

La fréquence de survie à	40 ans	des pers. de	30 ans	est	$\frac{63.082}{68.837}$
—	50	—	40	—	$\frac{56.057}{63.082}$
—	50	—	30	—	$\frac{56.057}{68.837}$

La dernière fraction est évidemment égale au produit des deux premières. Donc, la fréquence pouvant être considérée comme une mesure approchée de la proba-

bilité quand le nombre d'épreuves est très grand, si on appelle $_{10}p_{30}$; $_{10}p_{40}$; $_{20}p_{30}$, ces trois probabilités, on a, à la limite

$$_{20}p_{30} = {}_{10}p_{40} \times {}_{10}p_{30}.$$

Remarquons que le premier événement dont on cherche la probabilité consiste dans le concours des deux autres ; pour connaître le nombre des Français de 30 ans qui atteignent 50 ans, il suffit de savoir 1° combien de ces Français de 30 ans atteignent 40 ans ; 2° combien de ces Français de 40 ans atteignent 50 ans : mais, de ces deux derniers événements, le second suppose que le premier s'est produit. D'où l'énoncé suivant :

Lorsqu'un événement fortuit E consiste dans le concours de deux événements E_1, E_2, sa probabilité dans une certaine catégorie d'épreuves C est égale au produit de la probabilité de E_1 dans C par la probabilité de E_2 parmi les épreuves de C où E_2 s'est produit Donnons-en une démonstration générale. Considérons N épreuves de C. La répétition de l'événement E_1 étant m, le quotient $\frac{m}{N}$ sera voisin de la probabilité p_{E_1} de E_1. Soit q le nombre des épreuves où les événements E_1 et E_2 auront lieu ; $\frac{q}{N}$ sera voisin de la probabilité p_E de l'événement (E) résultant du concours de E_1 et de E_2. Or, on a évidemment

$$\frac{q}{N} = \frac{m}{N} \times \frac{q}{m}.$$

Que représente $\frac{q}{m}$: c'est le quotient ayant pour dénominateur la répétition de E_1, et pour numérateur la répétition non seulement de E_1 mais aussi de E_2. Donc $\frac{m}{q}$ est la probabilité de l'événement E_2 quand E_1 s'est produit, qu'on peut représenter par p'_{E_2} pour la distinguer de la probabilité pure et simple p_{E_2} de E_2. On a donc bien

$$p_E = p_{E_1} \times p'_{E_2}.$$

Le raisonnement est le même, que les événements E_1 et E_2 soient successifs ou simultanés.

Exemple. — Les tables d'assurance donnent les taux annuels de mortalité. Soit q_x la probabilité pour une personne d'âge x de mourir avant 1 année, p_x est la probabilité complémentaire, ou probabilité de survie au bout d'une année. On a :

$$p_x + q_x = 1.$$

Soit à calculer la probabilité ${}_np_x$ de survie au bout d'un certain nombre d'années n pour des personnes d'âge x. Pour qu'une personne d'âge x survive à $x+n$, il faut le concours des événements suivants : qu'elle survive à $x+1$; à $x+2$; etc. ; à $n+x$.

${}_np_x$ est donc égale au produit de la probabilité pour qu'une personne d'âge x survive à $x+1$ par la probabilité pour qu'une personne ayant atteint $x+1$ survive à $x+2$, et ainsi de suite.

$${}_np_x = p_x \, p_{x+1} \ldots p_{x+n-1}.$$

Événements indépendants. — Un cas particulier important du théorème est celui où les deux événements E_1, E_2 sont *indépendants*. Alors la probabilité de E_2 sera la même, que E_1 se soit produit ou non. On aura $p_{E_2} = p'_{E_2}$, d'où $p_E = p_{E_1} \times p_{E_2}$.

Exemple. — Le tableau ci-dessous est tiré d'une statistique anglaise[1] qui donne les âges des hommes mariés et des femmes mariées (sur plus de cinq millions de couples).

Fréquence des Anglaises mariées de 30 à 35 ans, parmi celles dont le mari a l'âge ci-dessous indiqué :

AGE DU MARI	FRÉQUENCE	OU SUR MILLE
20 à 25 ans.	0,017	17
25 à 30 —	0,122	122
30 à 35 —	0,503	503
35 à 40 —	0,317	317
40 à 45 —	0,101	101
45 à 50 —	0,034	34
50 à 55 —	0,007	17
55 à 60 —	0,008	8
Plus de 60	0,004	4

On a en outre cherché quelle était la fréquence des femmes mariées qui ont de 30 à 35 ans parmi toutes les femmes mariées, et on a trouvé 0,161 (soit 161 sur 1.000).

1. Census 1901, Summary tables, p. 182.

Considérons l'événement E_2 consistant pour une femme mariée à avoir entre 30 et 35 ans.

Considérons l'événement E_1 consistant pour son mari à avoir entre 20 et 25 ans.

Si l'événement E_2 était indépendant de E_1, cela signifierait que la fréquence des femmes mariées entre 30 et 35 ans reste sensiblement la même, quel que soit l'âge du mari. Sinon, la probabilité de E_2 serait différente (ou pourrait l'être) suivant que E_1 a ou n'a pas lieu.

La probabilité pour un couple que le mari ait de 20 à 25 ans et la femme de 30 à 35 ans est donc : $p_{E_1}\, p'_{E_2} = p_{E_1} \times 0{,}017$[1], et non pas $p_{E_1} \times p_{E_2} = p_{E_1} \times 0{,}161$, qui donnerait un nombre 10 fois trop fort. Les deux événements, dans ce cas, ne sont donc pas indépendants.

Si, au contraire, je prends au hasard deux conscrits, la probabilité P pour qu'ils aient tous deux une taille comprise entre $1^m{,}54$ et $1^m{,}62$ résultera de la probabilité P_1 de deux événements indépendants. $P_1 = 0{,}262$. On aura $P = P_1 \times P_1 = 0{,}262^2 = 0{,}069$.

Ici nous avons appliqué le théorème des probabilités composées sans restrictions[2].

Généralisation au cas de plus de deux événe-

1. Il nous arrivera souvent d'égaler approximativement les *valeurs numériques* de la fréquence et de la probabilité d'un même événement. Il ne saurait en résulter aucune confusion entre les notions distinctes de fréquence et de probabilité.

2. Il n'en serait pas de même si nous prenions par exemple deux frères : leurs tailles ne seraient pas indépendantes l'une de l'autre.

ments. — Le théorème des probabilités composées peut se généraliser, et on démontre comme plus haut qu'il s'applique aux cas où il y a plus de deux événements, pourvu que ce nombre soit fini.

Application. — Considérons des triplets formés de trois personnes prises au hasard, l'une parmi celles de 20 ans, l'autre parmi celles de 25 ans, la troisième parmi celles de 40 ans. Cherchons quelle est la probabilité P_n pour que ce groupe soit encore en vie au bout de n années. Les tables de mortalité ne donnent pas directement $_np_x$, c'est-à dire la probabilité pour qu'une personne d'âge x soit vivante au bout de n années, mais on y trouve les nombres l_{x+n} des personnes d'âge $x+n$ qui survivent en moyenne à cet âge parmi un groupe de l_x personnes d'âge x.

Donc comme on l'a vu

$$_np_x = \frac{l_{x+n}}{l_x}.$$

On aura, par le théorème des probabilités composées :

$$P_n = \frac{l_{20+n}}{l_{20}} \times \frac{l_{25+n}}{l_{25}} \times \frac{l_{40+n}}{l_{40}}.$$

Les quantités l_x varient d'une manière sensiblement continue : lorsque x croît de 20 à 40, $\frac{l_{x+n}}{l_x}$ passe donc par les trois valeurs indiquées, et sera compris entre la plus petite et la plus grande. Si, par exemple, les trois facteurs de P_n vont en décroissant les nombres $\left(\frac{l_{20+n}}{l_{20}}\right)^3$ et $\left(\frac{l_{40+n}}{l_{40}}\right)^3$ comprennent P_n ; et $\left(\frac{l_x+n}{l_x}\right)^3$ en

décroissant de $\left(\frac{l_{20+n}}{l_{20}}\right)^3$ à $\left(\frac{l_{40+n}}{l_{40}}\right)^3$ passera par la valeur de P_n. Il existera donc un âge x tel que $P_n = \left(\frac{l_{x+n}}{l_x}\right)^3$. Cet âge x est-il indépendant de la durée n?

Supposons le problème résolu affirmativement. Cela signifierait que la probabilité P_n est la même que la probabilité de survie d'un groupe de trois personnes d'âge x. On aurait ainsi substitué un groupe homogène (personnes du même âge) à un groupe hétérogène, ce qui permettrait de simplifier les calculs [1].

Faisons maintenant le calcul pour diverses durées : $n = 1$; $n = 5$; $n = 10$; $n = 20$; $n = 30$; en nous basant sur les tables de mortalité masculine de la France en 1901. Le problème est différent pour chaque valeur de n de sorte qu'il y a lieu de distinguer par des indices les différentes valeurs de x : x_1, x_5, x_{10}, x_{20}, x_{30}.

On aura :

$$\frac{l_{x+n}}{l_x} = \sqrt[3]{\frac{l_{20+n}}{l_{20}} \times \frac{l_{25+n}}{l_{25}} \times \frac{l_{40+n}}{l_{40}}}.$$

On trouve dans les tables les valeurs de l_{21}, l_{25},... l_{70} qui permettent de calculer les valeurs du second membre pour $n = 1$, 5,... soit, en appelant S_1, S_5,... les probabilités de survie du triplet homogène au bout de 1 année, 5 années, etc. :

$$S_1 = 0{,}991 \qquad S_5 = 0{,}955 \qquad S_{10} = 0{,}907$$
$$S_{20} = 0{,}788 \qquad S_{30} = 0{,}611.$$

1. Dans la pratique des assurances on arrive ainsi à éviter l'emploi de nombreuses tables à doubles, triples,... entrées.

Or on a :

PROBABILITÉ $\frac{l_{x+n}}{l_x}$ de vivre au bout de $n =$	1	5	10	20	30 années.
Pour des personnes âgées respectivement de $x =$ 31 ans . .	0,992	0,957	0,908	0,788	0,624
32 ans . .	0,992	0,956	0,907	0,781	0,608
33 ans . .	0,991	0,954	0,902	0,773	0,591

où les nombres ont été calculés à la règle à calcul (la dernière décimale est douteuse). Les nombres l_x et l_{x+n} ont été tirés des mêmes tables que les nombres précédents S_1, S_5,... Si on écrit au-dessous de ce tableau dans les colonnes $n = 1$; 5 ; 10 ; 20 ; 30 ; les probabilités respectives $\frac{l_{x_1+1}}{l_{x_1}} = S_1$,... calculées précédemment, on aura la quatrième ligne

$\frac{l_{x_n+n}}{l_{x_n}} =$	0,991	0,955	0,907	0,788	0,611

On voit qu'on serait amené à prendre

$$x_1 = 32 \text{ ou } 33\,;\ x_5 = 32 \text{ ou } 33\,;\ x_{10} = 32\,;\ x_{20} = 31\,;\ x_{30} = 31 \text{ ou } 32\text{ [1]}.$$

C'est-à-dire que la probabilité de survie d'un triplet

1. En partant des taux *bruts* de mortalité qui ont servi à établir la table des assurés français dite A. F., on trouverait de même $x_1 = 35$; $x_5 = 33$; $x_{10} = x_{20} = x_{30} = 32$.

de trois personnes âgées de 20, 25 et 40 ans est la même que celle d'un triplet de trois personnes âgées d'environ 32 ans, qu'il s'agisse d'une survie d'un an, de 5 ans, de 10 ans ou de 20 ans (loi de vieillissement uniforme).

Si on considère un plus grand nombre de personnes, il faudrait d'énormes calculs pour vérifier la loi dans sa généralité, savoir que :

Si on considère un groupe de personnes d'âges différents $X_1, X_2, \dots X_n$, on pourra toujours trouver un âge intermédiaire X, tel qu'un groupe de personnes d'âge X et en nombre égal à celui des personnes du du groupe donné ait la même probabilité de survie au bout d'une période quelconque.

C'est l'expérience des compagnies d'assurances qui avait mis sur la trace de cette loi : on avait remarqué, toutes les fois que la constatation avait pu être faite, qu'on pourrait remplacer dans les conditions précédentes un groupe de têtes d'âges différents par un groupe de têtes d'un âge intermédiaire (bien entendu, comme le montre l'exemple numérique ci-dessus, cet âge n'est pas nécessairement la moyenne arithmétique des différents âges).

On a réussi à remplacer chaque table de mortalité, tout au moins au-dessus de 20 ans, par une formule proposée par deux actuaires, Gompertz et Makeham, qui représente avec une bonne approximation par une expression mathématique la variation de l en fonction de x. Au moyen de cette formule, on démontre alors très facilement la loi de vieillissement uniforme dans

toute sa *généralité* et on détermine facilement X_q en fonction de $X_1, \ldots X_r$.

On a ici un exemple très frappant de l'utilité que présentent certaines notions mathématiques compliquées dans une question d'un intérêt pratique : ici, pour le fonctionnement d'une Compagnie d'assurances. Et, de même, dans le calcul numérique qui nous a conduit à cette loi, nous nous sommes appuyés principalement sur le théorème des probabilités composées, ce qui nous montre que ce théorème n'a pas un simple intérêt théorique.

Généralisation des deux théorèmes fondamentaux. — Considérons deux urnes A et A', contenant, la première, b blanches, n noires, la deuxième, b' blanches, n' noires, toutes ces boules étant supposées identiques. On se propose de chercher la probabilité P d'extraire une boule blanche de l'une des urnes A, A', l'urne où l'on doit tirer étant désignée par le sort. Par exemple on jette un dé, et l'on convient que s'il donne 6, on tirera dans A, s'il donne 5 ou 4, dans A', et qu'en tout autre cas on s'abstiendra.

Soit un grand nombre d'épreuves N. Cherchons la fréquence de l'événement favorable (consistant à extraire une blanche). Quelles seront ces épreuves ? Chacune consiste en l'ensemble de deux gestes : 1° On jette un dé et on note le point obtenu ; 2° on tire une boule de l'urne désignée, s'il y a lieu. Après chaque épreuve on remet la boule tirée dans son urne. Si le nombre

d'épreuves N est grand, il y en a un certain nombre m où le dé amène 6, q où il amène 4 ou 5, r où il amène 3, 2, 1. Sur les m épreuves, on tire de A un certain nombre de fois β une boule blanche. Sur les q épreuves, on tire de A' un nombre de fois β' une boule blanche.

La fréquence f de l'événement cherché est $\frac{\beta+\beta'}{N}$; si N est très grand, $\frac{\beta+\beta'}{N}$ sera généralement voisin de la probabilité cherchée.

Or on peut écrire

$$\beta = \frac{\beta}{m} \times m \qquad \beta' = \frac{\beta'}{q} \times q$$

pour mettre en évidence les fréquences secondaires.

Donc

$$P = \frac{\frac{\beta}{m} \times m + \frac{\beta'}{q} \times q}{N} = \frac{\beta}{m} \times \frac{m}{N} + \frac{\beta'}{q} \times \frac{N}{q}.$$

Considérons $\frac{m}{N}$. C'est la fréquence de l'événement consistant à obtenir le point 6 lorsqu'on jette un dé ; elle est d'environ $\frac{1}{6}$. $\frac{q}{N}$ est voisin de la probabilité du point 5 ou 4 dont chacun a une probabilité de $\frac{1}{6}$. Donc

$$\frac{q}{N} = \frac{2}{6} \text{ ou } \frac{1}{3}.$$

On a :

$$P = \frac{1}{6} \times \frac{\beta}{m} + \frac{1}{3} \frac{\beta'}{q}.$$

On voit que P est la somme du produit de la probabilité de choisir l'urne A par la probabilité d'en tirer une boule blanche et du produit de la probabilité de choisir l'urne A' par la probabilité d'en tirer une boule blanche. Ce résultat, auquel nous sommes arrivés par un calcul direct, sur un exemple concret, nous l'aurions obtenu de façon abstraite et indirecte, mais plus rapidement, en utilisant nos deux théorèmes des probabilités totales et composées.

Plaçons-nous dans le cas général. Soit un événement fortuit E. Je suppose que E puisse se produire de plusieurs manières E_1, E_2... E_n, qui s'excluent mutuellement, et que chacune de ces manières soit conditionnée par un événement fortuit A_1, A_2... A_n. Cherchons la probabilité p de E, connaissant les probabilités p_1 de E_1 quand A_1 s'est produit, p_2 de E_2 quand A_2 s'est produit, etc., et aussi ϖ_1, probabilité de A_1, ϖ_2, probabilité de A_2, etc. (dans l'exemple précédent, E est le tirage d'une blanche de l'une des urnes, chaque manière E_1, etc., consistant à la tirer d'une urne déterminée, les événements A_1, A_2, etc., étant les points amenés par le dé).

D'après le théorème des probabilités totales, p est la somme des probabilités correspondant aux diverses manières dont l'événement peut se produire. Mais chacune de ces manières est conditionnée elle-même par un événement, représente donc le concours de deux événements dont chacun a une probabilité déterminée. Sa probabilité est donc, d'après le théorème des probabi-

lités composées, le produit de ces deux probabilités secondaires : $p_1 \pi_1$, $p_2 \pi_2$, ...

Donc :

$$p = p_1 \times \pi_1 + p_2 \times \pi_2 + \dots$$

Ce qui généralise à la fois les théorèmes des probabilités totales et composées.

Application de ces théorèmes à la solution de quelques problèmes.

1° Quelle est la probabilité pour que, sur trois têtes d'âges a, b, c, il y en ait au moins une qui survive au bout de n années ? Ce problème nous donnera l'occasion de montrer comment on utilise les probabilités dites *complémentaires*, ou de l'événement contraire. Soit p la probabilité cherchée ; q, la probabilité complémentaire, est égale à $1 - p$. C'est la probabilité pour qu'au bout de n années les trois personnes soient mortes. Appelons ${}_nq_a$, ${}_nq_b$, ${}_nq_c$ la probabilité de mourir pour chacune des têtes. Comme ces probabilités sont indépendantes, q est égal au produit de ces trois quantités. Et la probabilité cherchée, calculée ainsi par le détour de la probabilité complémentaire, est

$$p = 1 - q = 1 - ({}_nq_a)\,({}_nq_b)\,({}_nq_c).$$

Les tables de survie permettent d'ailleurs de calculer facilement le second membre ; donnant le nombre l_{a+n} des personnes qui survivent au bout de n années

d'un groupe de l_a personnes d'âge a, elles fournissent

$$_nq_a = \frac{l_a - l_{a+n}}{l_a} \text{ et de même } {}_nq_b = \frac{l_b - l_{b+n}}{l_b}, \; {}_nq_c = \frac{l_c - l_{c+n}}{l_c}$$

2° D'après une vieille croyance populaire, si treize personnes se mettent ensemble à table, l'une d'elles au moins mourra avant une année. On pourrait préciser, en disant : la probabilité pour que, sur treize personnes prises au hasard, l'une d'elles au moins meure avant une année est voisine de la probabilité de la certitude, c'est-à-dire de l'unité. Est-ce exact?

Remarquons d'abord que la probabilité pour que, dans un groupe de personnes déterminé, l'une au moins meure avant une année est évidemment d'autant plus grande que ce groupe est plus nombreux. Il s'ensuit que, si treize personnes s'en adjoignent une quatorzième, la probabilité de l'événement redouté ne fera qu'augmenter.

Recherchons cependant si l'éventualité en question est réellement vraisemblable, et évaluons à cette fin sa probabilité. On pourrait appliquer ici une méthode expérimentale, c'est-à-dire observer un grand nombre de groupes de 13 personnes pendant une année, et noter la fréquence de ceux de ces groupes où un décès au moins a été constaté. Il est évident qu'une telle observation serait plus compliquée que celle qui donnerait la probabilité de décès d'*une* personne en une année. De plus tout serait à refaire, si l'on considérait un groupe de 14 personnes par exemple au lieu de 13.

La méthode qui nous a fourni la dernière formule (page 34) nous permet de faire l'économie de ces observations.

Supposons par exemple que les âges des treize personnes soient respectivement :

20 ; 25 ; 30 ; 35 ; 40 ; 45 ; 50 ; 55 ; 60 ; 65 ; 70 ; 75 ; 80 ans,

et appelons p_{20}, p_{25},.... p_{80} leurs probabilités respectives de ne pas mourir dans l'année.

Du raisonnement fait plus haut il ressort que la probabilité ϖ pour que l'une au moins de ces treize personnes meure dans l'année est :

$$\varpi = 1 - p_{20} \times p_{25} \times \dots p_{80}.$$

Si par exemple on prend pour p_{20}, p_{25},.... les valeurs données par la statistique générale de la France d'après le recensement de 1901, on aura :

$$\begin{aligned}\varpi = 1 - (&0{,}99301 \times 0{,}99248 \times 0{,}99214 \times 0{,}99058 \\ &\times 0{,}98896 \times 0{,}98637 \times 0{,}98299 \times 0{,}97847 \times 0{,}96916 \\ &\times 0{,}95570 \times 0{,}93168 \times 0{,}89130 \times 0{,}83220) = 0{,}415.\end{aligned}$$

On voit que cette probabilité n'est pas très grande. Ce n'est, par exemple, même pas la probabilité d'amener face dans le jeu de pile ou face.

Nous avons choisi arbitrairement les âges des treize personnes. On peut se demander si le résultat obtenu serait très différent pour un choix autre. Pour nous en rendre compte, cherchons quel est l'âge minimum que devraient avoir les treize personnes, supposées cette

fois toutes du même âge, pour que la probabilité ϖ soit très grande, c'est-à-dire voisine de l'unité. Or on aurait dans ce cas, en remplaçant p_{20}, p_{25}, ... p_{x0} par p_x :

$$\varpi = 1 - (p_x)^{13} \quad \text{d'où} \quad (p_x)^{13} = 1 - \varpi.$$

Pour que la probabilité $1 - \varpi$ qu'aucune des 13 personnes d'âge x ne meure avant une année soit au moins aussi petite que, par exemple, celle de former le mot : miracle, en tirant successivement 7 caractères d'imprimerie au hasard, il faut que (voir exemple V, p. 39) :

$$(p_x)^{13} < \frac{1}{25 \times 24 \times 23 \times 22 \times 21 \times 20 \times 19}$$

$$\frac{1}{2.422.728.000}.$$

Le calcul montre que p_x doit être plus petit que 0,191, et par conséquent, en se reportant à la table de survie indiquée, x doit être supérieur à 103. L'âge commun des treize personnes devrait être d'au moins 103 ans. En d'autres termes, il arrivera aussi rarement que, de treize personnes toutes âgées d'au moins 103 ans, aucune ne meure avant une année, qu'il arrivera que, tirant au hasard sept caractères d'imprimerie successivement, on forme le mot : miracle.

Mais soyons moins exigeants. Contentons-nous d'une probabilité un peu moins voisine de la certitude. Pour que la certitude de voir une des treize personnes d'âge x mourir avant une année soit aussi grande que, par exemple, celle de ne pas amener trois fois de suite le

point six au jeu de dés, il faudrait[1] que $(p_x)^{13} < \frac{1}{6^3}$, d'où : $p_x < 0,661$, et alors $x > 94$. L'âge obtenu, soit 95 ans, reste encore assez élevé. Et cependant il s'agit d'une certitude beaucoup moins grande que dans le cas précédent.

Problèmes tirés des jeux de hasard.

I. Quelle est la probabilité de tirer un roi d'un jeu de 32 cartes ? L'expérience montre que la probabilité de tirer chacune des cartes est la même, elle est donc égale à $\frac{1}{32}$. La probabilité de tirer un roi est la somme des probabilités des événements s'excluant mutuellement qui consistent, le premier, à tirer un roi de cœur, le second, un roi de carreau...... Elle est donc égale à

$$\frac{1}{32} + \frac{1}{32} + \frac{1}{32} + \frac{1}{32}, \text{ soit à } \frac{4}{32} \text{ ou } \frac{1}{8}$$

II. Considérons deux jeux de 32 cartes. On tire une carte du premier, une du second. Quelle est la probabilité de tirer deux rois ? Soit E l'acte de tirer deux rois : il résulte de deux événements, E_1 et E_2, indépendants. Le théorème des probabilités composées s'applique donc sans condition. La probabilité cherchée est

$$\frac{1}{8} \times \frac{1}{8} = \frac{1}{64}.$$

1. Nous emploierons fréquemment la notation $a < b$ pour : a plus petit que b.

III. Considérons un seul jeu (au lieu de deux), d'où je tire deux cartes. Quelle est la probabilité que j'en tire deux rois ? Ici, l'événement composé E résulte de deux événements E_1, E_2, qui ne sont pas indépendants. La probabilité de E est donc égale à la probabilité de E_1 multipliée par la probabilité de E_2 quand E_1 s'est produit $= \frac{1}{8} \times \frac{3}{31}$ (puisqu'après que le premier roi est tiré il reste 31 cartes et 3 rois).

Remarquons que si l'on avait tiré la seconde carte après avoir remis la première dans le jeu, on aurait le même résultat que dans la question II.

IV. Une urne contient n boules noires, b boules blanches, identiques sauf la couleur. L'expérience confirme ce que notre observation inconsciente nous conduit à deviner : que les événements consistant à tirer une boule déterminée ou une autre boule déterminée sont également possibles. La probabilité de tirer une boule blanche est donc $\frac{b}{b+n}$, puisqu'on peut alors distinguer $b + n$ cas également possibles dont b sont favorables.

V. On considère un jeu de 25 cartes sur chacune desquelles on a écrit respectivement les 25 lettres de l'alphabet. On cherche la probabilité P pour qu'en tirant successivement un certain nombre de cartes on forme un mot donné, par exemple pour qu'en tirant sept cartes on forme le mot : miracle.

La probabilité, au premier tirage, de tirer m est $\frac{1}{25}$;

la probabilité de tirer i du reste des cartes est $\frac{1}{24}$, de tirer r, $\frac{1}{23}$, et ainsi de suite, jusqu'à la probabilité de tirer e qui est $\frac{1}{19}$. Donc $P = \frac{1}{25 \times 24 \times \cdots \times 19}$. Ce résultat est évidemment inférieur à $\frac{1}{20^7}$, et *a fortiori* inférieur à un milliardième. Remarque : le même calcul s'appliquerait à tout mot de 7 lettres, pourvu que celles-ci soient distinctes.)

On peut chercher la probabilité P' pour qu'en tirant d'une boîte d'imprimerie sept caractères simultanément, on obtienne ceux du mot miracle dans un ordre quelconque. Une première méthode sera la suivante : on remarque que, si l'épreuve est favorable, on tire une lettre quelconque du mot miracle, puis, l'ayant enlevée, on tire une des six lettres du mot qui restent, etc. La probabilité P_1 de tirer la première lettre quelconque du mot, est égale à la probabilité de tirer une blanche d'une urne qui contient 25 boules, dont 7 blanches : $= \frac{7}{25}$. La probabilité de tirer la seconde lettre quelconque du même mot, $P_2 = \frac{6}{24}$, et ainsi de suite. Donc $P' = \frac{7}{25} \times \frac{6}{24} \times \cdots \times \frac{1}{19} = P \times$ par le produit des sept premiers nombres entiers. — Une autre méthode consiste à se servir de la précédente en remarquant que la valeur de P serait la même, quel que soit le groupe de 7 cartes qu'il s'agisse de tirer du lot de 25. Autrement dit les tirages de ces différents groupes sont également vraisem-

blables. Ils représentent le nombre des cas possibles N, et P′ est la probabilité de l'une d'elles $= \frac{1}{N}$. Mais N = le nombre de combinaisons de 25 cartes 7 à 7, nombre que nous pouvons désigner par C^7_{25}.

D'où

$$P' = \frac{1}{C^7_{25}}.$$

Nous sommes donc ramenés à un problème de pure algèbre : calculer C^7_{25}. Inversement la comparaison des deux valeurs de P′ nous fournit la solution de ce problème d'algèbre.

$$C^7_{25} = \frac{25 \times \ldots \times 19}{1 \times 2 \times \ldots \times 7}.$$

Remarquons que, pour que cette démonstration fût exacte, il faudrait prouver expérimentalement que les tirages des différents groupes étaient également probables. Mais le théorème des probabilités composées permet de vérifier ici cette induction, c'est-à-dire d'en ramener la preuve expérimentale à celle de l'égale probabilité des tirages d'une seule carte déterminée.

Plus généralement, considérons n cartes sur lesquelles sont représentés n objets distincts. Calculons la probabilité pour qu'en tirant à la fois q de ces n cartes ($q \leq n$) on obtienne des cartes représentant q objets déterminés d'avance, dans un ordre quelconque.

1re *méthode :*

$$P' = \frac{q}{n} \times \frac{q-1}{n-1} \times \ldots \times \frac{1}{n-q+1}.$$

2^e *méthode :* $P' = \frac{1}{C_n^q}$ en appelant C_n^q le nombre de combinaisons de n objets q à q. La comparaison de ces deux égalités nous fournit indirectement *une formule qui nous sera souvent utile par la suite* (et nous dispensera de supposer connue la théorie algébrique appelée analyse combinatoire) :

$$(1) \qquad C_n^q = \frac{n(n-1)\ldots(n-q+1)}{1 \times 2 \ldots q}.$$

VI[1]. Une urne contient b boules blanches, n boules noires. On en tire à la fois q boules. Quelle est la probabilité pour que ces q boules soient toutes blanches?

Naturellement il y a impossibilité, c'est-à-dire probabilité nulle, si q est supérieur à b. Supposons donc que q est égal au plus à b.

On a vu (question IV) que la probabilité, en tirant une boule, qu'elle soit blanche est $\frac{b}{b+n}$.

L'événement considéré consiste dans le concours de q événements dont chacun est le tirage d'une boule blanche. La probabilité correspondante x est donc le produit des probabilités : pour que, sur les $b+n$ boules, on en tire une des b blanches au premier tirage, pour que, ce tirage étant fait, sur les $b+n-1$ boules restantes on en tire une des $b-1$ blanches, ... jusqu'au q^e tirage inclus. Donc

$$x = \frac{b}{b+n} \times \frac{b-1}{b+n-1} \times \ldots \times \frac{b-q+1}{b+n-q+1}.$$

1. On pourra laisser de côté, à première lecture, la suite de ce chapitre.

Par exemple : il y a dans l'urne 10 blanches et 10 noires ; on en tire 5 boules ; la probabilité pour qu'elles soient toutes blanches est :

$$\frac{10}{20} \times \frac{9}{19} \times \frac{8}{18} \times \frac{7}{17} \times \frac{6}{16} = \frac{21}{1292} < \frac{1}{60}.$$

VII. Résolvons le même problème en supposant que les boules soient tirées une à une, et qu'après en avoir tiré une et avoir noté sa couleur, on la replace dans l'urne avant de tirer la suivante.

Alors la probabilité de tirer une blanche reste la même $\left(\frac{b}{b+n}\right)$ à chacun des q coups. La probabilité cherchée est alors : $\left(\frac{b}{b+n}\right)^q$. Dans l'exemple considéré, elle serait $\left(\frac{10}{20}\right)^5 = \frac{1}{32}$, c'est-à-dire presque deux fois plus grande.

VIII. *Plus généralement*, une urne contient b boules blanches, n noires. On en tire à la fois q boules. On demande la probabilité P pour que β d'entre ces q boules soient blanches.

Cherchons d'abord quelle est la probabilité p pour que β des q boules soient blanches, les rangs des β tirages où la boule tirée est blanche étant déterminés d'avance. Alors soit x_r le nombre des celles des boules restant dans l'urne avant le tirage d'ordre r et qui son de la couleur fixée d'avance pour ce tirage. La probabi lité p sera :

$$p = \frac{x_1}{b+n} \times \frac{x_2}{b+n-1} \times \dots \frac{x_q}{b+n-q+1}.$$

Dans les nombres x_1 ..... x_q sont compris les nombres des boules de la couleur qu'on désire tirer. Donc β des nombres x se rapportent aux tirages des boules blanches, et sont égaux respectivement à b, $b - 1$, $b - 2$, ... $b - \beta + 1$, et de même $q - \beta$ des autres nombres x sont égaux respectivement à n, $n - 1$, ... $n - q + \beta + 1$ (pour les noires). Les nombres x_1, ... x_q sont formés des nombres b, $b - 1$ ..... $b - \beta + 1$, n, $n - 1$ ..... $n - q + \beta + 1$ rangés dans un certain ordre (quelconque). Donc :

$$p = \frac{b(b-1)\ldots(b-\beta+1)\,n(n-1)\ldots(n-q+\beta+1)}{(b+n)\ldots(b+n-q+1)}.$$

Si $\beta = 0$, p est la probabilité de tirer q noires successivement ; d'après le théorème des probabilités composées, c'est donc

$$\frac{n}{b+n} \cdot \frac{n-1}{b+n-1} \cdots \frac{n-q+1}{b+n-q+1}.$$

On voit alors que p ne dépend pas des rangs où se trouvent, dans les q tirages les tirages d'une boule blanche. On aura : $P = kp$. A quoi est égal k ? C'est le nombre de dispositions différentes des q lettres B .... B N .... N. Ces dispositions diverses se distinguent par les β rangs des β lettres B, ces rangs étant pris dans 1 ..... q. Donc k est égal au nombre des combinaisons des q premiers rangs β à β ; $k = C_q^\beta$; bien entendu $k = 1$ si $\beta = 0$.

Donc si $\beta = 0$, $P = p =$ la valeur de l'expression précédente, quand $\beta = 0$, soit :

$$P = \frac{n(n-1)\ldots(n-q+1)}{(b+n)(b+n-1)\ldots(b+n-q+1)}.$$

Sinon :

$$P = \frac{q(q-1)\ldots(q-\beta+1)}{\beta\ldots 2.1} \, \frac{b(b-1)\ldots(b-\beta+1)n\ldots(n-q+\beta+1)}{(b+n)\ldots(b+n-q+1)}$$

d'après la formule (1).

Factorielles. — A propos de ces formules, signalons une *notation* qui permet soit d'abréger les formules, soit (comme on le verra ici à titre d'exemple) d'en faciliter les transformations.

On désigne par $n\,!$ (et on prononce « factorielle n ») le produit des n premiers nombres entiers :

(2) $$n! = 1 \times 2 \times 3 \times \ldots\ldots \times (n-1) \times n$$

par exemple :

$$5\,! = 1 \times 2 \times 3 \times 4 \times 5.$$

Le nombre des combinaisons de n objets q à q que nous avons désigné par C_n^q et que nous avons évalué à

$$\frac{n(n-1)\ldots(n-q+1)}{1 \times 2 \times \ldots \times q}$$

peut aussi s'écrire

$$\frac{n(n-1)\ldots(n-q+1)(n-q)\ldots 2 \times 1}{1 \times 2 \times \ldots \times q \times (n-q)\ldots \times 2 \times 1}$$

et par suite :

(3) $$C_n^q = \frac{n\,!}{q\,!\,(n-q)\,!}.$$

Alors on voit que la formule qui résout la dernière question

$$P =$$

$$\frac{q(q-1)\ldots(q-\beta+1)}{\beta\ (\beta-1)\ldots 2\ .\ 1}\ \frac{b(b-1)\ldots(b-\beta+1)n\ldots(n-q+\beta+1)}{(b+n)\ldots(b+n-q+1)}$$

peut s'écrire :

ou bien en la mettant sous la forme qui lui a donné naissance

$$P = C_q^\beta \times p \quad \text{avec}$$

$$p = \frac{b(b-1)\ldots(b-\beta+1)\,n\ldots(n-q+\beta+1)}{(b+n)\ldots(b+n-q+1)}$$

$$= \frac{\dfrac{b\,!}{(b-\beta)\,!}\ \dfrac{n\,!}{(n-q+\beta)\,!}}{\dfrac{(b+n)\,!}{(b+n-q)\,!}}$$

$$= \frac{(b+n-q)\,!}{(b-\beta)\,!\,(n-q+\beta)\,!} \times \frac{1}{\dfrac{(b+n)\,!}{b\,!\,n\,!}} = \frac{C_{b+n-q}^{b-\beta}}{C_{b+n}^{b}}$$

d'où la forme abrégée

$$P = \frac{C_q^\beta\ C_{b+n-q}^{b-\beta}}{C_{b+n}^{b}}\ ;$$

ou bien en procédant de la même manière sur l'expression complète de P

$$P = \frac{\dfrac{q\,!}{(q-\beta)\,!}}{\beta\,!}\ \frac{\dfrac{b\,!}{(b-\beta)\,!}\ \dfrac{n\,!}{(n-q+\beta)\,!}}{\dfrac{(b+n)\,!}{(b+n-q)\,!}}$$

$$= \frac{b\,!}{\beta\,!\,(b-\beta)\,!}\ \frac{n\,!}{(q-\beta)\,!\,(n-q+\beta)\,!}\ \frac{1}{\dfrac{(b+n)\,!}{q\,!\,(b+n-q)\,!}}$$

d'où une seconde forme abrégée

$$P = \frac{C_b^\beta \, C_n^{q-\beta}}{C_{b+n}^q}$$

Généralisation des théorèmes des probabilités totales et composées au cas d'un ensemble d'événements compatibles et dépendants. — D'après le théorème des probabilités totales, si deux événements E_1, E_2 sont incompatibles, la probabilité Q_{12} pour que l'un ou l'autre se produise est égale à la somme des probabilités q_1, q_2, de chacun de ces événements. Si ces événements sont compatibles, on a, comme nous allons le voir, $Q_{12} \leqq q_1 + q_2$. Mais on peut arriver à des formules plus précises. — Appelons q'_1 la probabilité de l'événement E_1 quand E_2 n'a pas lieu, et q'_2 celle de E_2 quand E_1 n'a pas lieu. Appelons enfin q_{12} la probabilité pour que E_1, E_2 aient lieu à la fois, et Q'_{12} la probabilité que E_1 ait lieu seul ou E_2 seul.

E_1 peut avoir lieu de deux façons, soit seul, soit avec E_2, et ces deux hypothèses sont incompatibles.

Donc la probabilité de E_1, ou $q_1 = q'_1 + q_{12}$. D'où

(4) $$q'_1 = q_1 - q_{12}.$$

Il peut arriver que ou bien E_1 (avec ou sans E_2), ou bien E_2 seul ait lieu. Donc

$$Q_{12} = q_1 + q'_2 = q_2 + q'_1.$$

D'où, en remplaçant q'_1 par sa valeur.

(5) $$Q_{12} = q_2 + q_1 - q_{12}.$$

Ainsi non seulement nous savons que $q_1 + q_2 - Q_{12}$ est positif, mais nous savons aussi interpréter sa valeur : q_{12}.

Considérons maintenant la probabilité (considérée comme unique), que E_1 ait lieu seul, ou E_2 seul, (événements incompatibles)

$$Q'_{12} = q'_1 + q'_2$$

ou, d'après la formule (4)

$$Q'_{12} = q_1 - q_{12} + q_2 - q_{12} = q_1 + q_2 - 2q_{12}. \tag{6}$$

Passons au cas de trois événements F_1, F_2, F_3, dont les probabilités respectives sont r_1, r_2, r_3 ; appelons r'_1 la probabilité pour que F_1 ait lieu sans F_2 ni F_3 ; et, de même r'_2 la probabilité pour que F_2 ait lieu sans F_1 ni F_3 ; et de même r'_3, pour que F_3 ait lieu sans F_1 ni F_2. Appelons r_{12} la probabilité pour que F_1 et F_2 aient lieu à la fois ; R_{12}, pour que F_1 ou F_2 aient lieu ; R'_{12}, pour que F_1 ait lieu et non F_2, ou que F_2 ait lieu et non F_1 ; enfin R_{123} pour que l'un au moins des trois événements F_1, F_2, F_3 se produisent, et r_{123} pour qu'ils aient lieu à la fois.

Si on considère comme événements E_1, E_2 les événements F_1 d'une part, et d'autre part celui qui consiste dans la réalisation de F_2 ou F_3, on aura comme précédemment :

$$q'_1 = q_1 - q_{12}; \quad Q_{12} = q_1 + q_2 - q_{12}; \quad Q'_{12} = q_1 + q_2 - 2q_{12}$$

qui donnent maintenant

$$r'_1 = r_1 - q_{12}; \quad R_{123} = r_1 + q_2 - q_{12}; \quad Q'_{12} = r_1 + q_2 - 2q_{12}.$$

De plus q_{12} est la probabilité pour qu'à la fois aient lieu F_1 et soit F_2, soit F_3 ; en appliquant une formule analogue à (5) on a :

$$q_2 = r_2 + r_3 - r_{23}$$
$$q_{12} = r_{12} + r_{13} - r_{123}.$$

D'où

$$R_{123} = r_1 + (r_2 + r_3 - r_{23}) - (r_{12} + r_{13} - r_{123})$$
$$(7) \qquad R_{123} = r_1 + r_2 + r_3 - (r_{23} + r_{12} + r_{13}) + r_{123}$$

Puis

$$(8) \qquad r'_1 = r_1 - (r_{12} + r_{13}) + r_{123}.$$

D'autre part soit R'_{123}, la probabilité pour qu'un seul des événements F_1, F_2, F_3 ait lieu. C'est donc la somme des probabilités

r'_1 pour que F_1 ait lieu et non F_2 ni F_3
r'_2 — F_2 — — F_1 —
r'_3 — F_3 — — — ni F_2

ou, d'après (8)

$$\begin{aligned} R'_{123} = r'_1 + r'_2 + r'_3 &= r_1 - (r_{12} + r_{13}) + r_{123} \\ &+ r_2 - (r_{23} + r_{21}) + r_{123} \\ &+ r_3 - (r_{31} + r_{32}) + r_{123} \end{aligned}$$
$$(9) \qquad R'_{123} = r_1 + r_2 + r_3 - 2\,(r_{12} + r_{23} + r_{31}) + 3\,r_{123}.$$

Formule générale pour un nombre quelconque d'événements. — On passera de même au cas de 4 et d'un plus grand nombre d'événements H_1, H_2, ... compatibles ou non.

Appelons p_1 la probabilité de H_1, p_2 celle de H_2, ... p'_1 la probabilité que H_1 ait lieu sans H_2, ni H_3, ni ..... ;

P la probabilité pour que l'un au moins, quel qu'il soit, de ces événements ait lieu ; P′ la probabilité pour qu'un seul, quel qu'il soit, de ces événements ait lieu. On a alors :

$$p'_1 = p_1 - \sum_h p_{1h} + \sum_{h,k} p_{1hk} - \sum_{h,k,l} p_{1hkl} + \dots$$
$$P' = \sum_i p_i - 2 \sum_{i,h} p_{ih} + 3 \sum_{i,h,k} p_{ihk} - \dots$$
$$P = \sum_i p_i - \sum_{i,h} p_{ih} + \sum_{i,h,k} p_{ihk} - \dots$$

en désignant par $\Sigma_i\, p_i$ la somme $p_1 + p_2 + \dots\dots + p_n$ des termes obtenus en remplaçant i par 1, 2 ... h ; par $\sum\limits_{ih} p_{ih}$ la somme $p_{12} + p_{13} + \dots + p_{1n} + p_{23} + p_{24} + \dots + p_{2n} + \dots + p_{nn-1}$ des termes obtenus en remplaçant i, h par deux termes *différents* de la suite 1, 2 ..... n, etc., et enfin par $p_{ih\,\dots\dots\,s}$ la probabilité pour que $H_i\ H_h \dots\dots H_s$ aient lieu à la fois

En employant une notation symbolique dont la signification est évidente, la dernière de ces formules, qui est due à Poincaré, s'écrit :

$$(H_1 + \dots + H_h) = \sum_i (H_i) - \sum_{i,h} (H_i . H_h) + \sum_{i,h,k} (H_i . H_h . H_k) - \dots$$

Inégalités relatives aux événements dépendants. — Tâchons d'obtenir quelques résultats qui permettent d'étendre le théorème des probabilités composées au cas des événements dépendants. Ce théorème fournit une égalité reliant $p_1 \dots\dots p_n$ et $p_{12\,\dots\dots\,n}$ quand $H_1 \dots\dots H_n$ sont indépendants. Dans le cas général, si l'égalité ne subsiste plus, on peut cependant écrire une

inégalité qui donne un renseignement sur $p_{12 \ldots n}$ quand on connaît *seulement* p_1, $p_2, \ldots p_n$. On a, d'après (5) $p_1 + p_2 - p_{12} =$ la probabilité pour que H_1 ou H_2 ait lieu.

Donc

$$0 \leqq p_1 + p_2 - p_{12} \leqq 1$$

Par suite

$$(10) \qquad \begin{aligned} &p_{12} \leqq p_1 + p_2 \\ &\qquad p_{12} \geqq p_1 + p_2 - 1 \end{aligned}$$ [1]

(La première inégalité n'est pas intéressante, car on a évidemment $p_{12} \leqq p_1$ et $p_{12} \leqq p_2$, inégalités qui sont plus précises).

Pour trois événements :

$$p_{123} \geqq p_3 + p_{12} - 1 \geqq p_1 + p_2 + p_3 - 2$$

et pour n événements :

$$p_{1 \cdots n} \geqq p_1 + \cdots + p_n - (n-1)$$

Cette inégalité peut rendre surtout service quand les événements $E_1 \ldots\ldots E_n$ sont presque certains.

Si en effet $1 - p_1, \ldots\ldots, 1 - p_n$ sont respectivement inférieurs à des nombres très petits $\varepsilon_1, \ldots\ldots \varepsilon_n$, on a

$$(11) \qquad \begin{aligned} &p_1 \geqq 1 - \varepsilon_1 \ldots p_n \geqq 1 - \varepsilon_n \\ &p_{12 \ldots n} \geqq 1 - (\varepsilon_1 + \ldots + \varepsilon_n) \end{aligned}$$

qui sera voisin de 1.

1. Pour abréger le langage nous employons le signe $\geqq$, qui se prononce « supérieur ou égal à ». De même le signe $<$ se prononce « inférieur à ». Par exemple $3 < 4$.

Ainsi donc, si les événements contraires à $E_1 \ldots\ldots E_n$ ont de très petites probabilités, l'événement contraire à leur réalisation simultanée aura aussi une très petite probabilité. De plus la formule (11) permet d'exprimer numériquement ce fait et cela est important. Si, en effet, une cinquantaine d'événements ont une probabilité $> 1 - \frac{1}{100}$, leur réalisation simultanée aura une probabilité $> 1 - \left(\frac{1}{100} \times 50\right) = \frac{1}{2}$, ce qui n'est pas très voisin de 1. Les affirmations précédentes ne seront donc valables que si les nombres des événements ne sont pas trop grands pour un ordre de grandeur donné des probabilités élementaires.

Supposons par exemple qu'on veuille fixer les conditions de recrutement des conscrits.

On sait que la fréquence d'une taille supérieure à a centimètres est 95 p. 100.

On sait que la fréquence d'un tour de poitrine supérieure à b centimètres est 90 p. 100.

On sait que la fréquence d'une vue inférieure à c dioptries est 75 p. 100.

Alors la proportion des conscrits vérifiant ces conditions sera $\geqq 1 - (0{,}05 + 0{,}10 + 0{,}25)$.

On voit tout de suite qu'il serait dangereux d'imposer un trop grand nombre de ces conditions (à moins qu'elles soient étroitement dépendantes), sinon on réduirait trop le nombre des conscrits admissibles.

EXERCICES

I. — Combien faut-il réunir de convives choisis au hasard mais tous du même âge, pour qu'il y ait au moins une chance sur deux qu'un au moins des convives meure avant une année. On fera le calcul : 1° quand l'âge commun des convives est 20 ans et en admettant qu'à cet âge le taux annuel de mortalité est 0,00699 ; 2° en substituant à ces deux nombres : 40 ans et 0,01104.

La solution mettra en évidence d'une autre manière, la conclusion à laquelle nous sommes déjà arrivés dans l'étude du problème de la page 35.

II. — On comptait en 1921 dans les trois départements du Bas-Rhin, du Haut-Rhin et de la Moselle 1.703 communes qui pouvaient être classées de la façon suivante d'après l'importance de leur population.

Nombre d'habitants.	Nombre des communes.
de 0 à 1.000	1.399
1.001 à 2.000	185
2.001 à 5.000	81
plus de 5.000	38

Supposons qu'on tire au sort les noms de quatre des maires de ces 1.703 communes. On demande de calculer, en se basant sur le recensement de 1921, la probabilité pour que la délégation ainsi constituée se trouve être représentative en ce sens qu'elle contienne un représentant de chacune des catégories de communes mentionnées ci-dessus.

CHAPITRE II

THEORIE DES PROBABILITÉS CONTINUES OU GÉOMÉTRIQUES

Probabilité dépendant de la position d'un point sur une droite, ou d'une variable. — Nous avons pu jusqu'ici distinguer nettement les événements favorables et défavorables, par exemple, au jeu de dé, à pile ou face, etc. Non seulement la différence entre les uns et les autres est bien marquée, mais encore, une fois qu'on les a définis, on ne trouve point d'intermédiaires entre eux. La façon dont ils se réalisent n'importe pas, à cet égard : que la pièce tombe pile à plat, ou après avoir roulé plus ou moins longtemps, pile reste aussi différent de l'événement contraire face. En d'autres termes on ne passe pas de l'un à l'autre de façon continue.

Mais des cas se présentent où il n'en est plus ainsi. On envisage par exemple la probabilité pour qu'un point mobile sur une droite soit situé sur un segment AB de cette droite. L'événement sera favorable lorsque M sera entre A et B, défavorable lorsqu'il sera en dehors. Mais suivant que M sera plus ou moins près de A ou de B, l'événement, favorable ou défavorable, se différenciera plus ou moins nettement de l'événement con-

traire. On dit, dans ce cas, qu'il s'agit d'une probabilité continue [1].

Nous allons examiner de près cet exemple qui est fondamental dans la théorie des probabilités continues. — On dit généralement, mais à tort selon nous, qu'il faut faire une convention, et s'entendre sur ce qu'on conviendra d'appeler la probabilité dans le cas des probabilités continues. En réalité, si la théorie des probabilités est tirée de l'expérience physique, il n'y doit rien entrer de conventionnel. Rien ne m'empêche d'admettre que la définition de la probabilité comme une grandeur physique, dont la fréquence est une mesure approchée, s'applique aussi bien aux probabilités continues qu'aux discontinues ; mais il me faut définir la catégorie d'épreuves envisagées (comme précédemment). Comment s'effectuera matériellement l'expérience ? Nous allons voir que, suivant les cas, les résultats peuvent être différents.

Exemples. — Par exemple, dans le problème précédent, on peut supposer que chaque épreuve consiste à lancer une aiguille sur un plan qui contient la droite, et à noter la position du point de rencontre de la droite

1. En réalité, dans les problèmes de probabilité discontinue, on pourrait considérer des termes intermédiaires. Par exemple, s'il s'agit d'amener un six, le dé se trouvera quelquefois dans une position d'équilibre où il est près d'osciller vers le six, mais aussi bien vers un autre point. Seulement il faudrait alors convenir de distinguer les cas non point seulement d'après le résultat final, mais d'après les positions successives du dé avant qu'il soit tombé.

avec l'aiguille (on écartera les cas où l'aiguille tombe en dehors de la droite). Pour plus de précision, on pourra supposer que l'aiguille est toujours lancée de façon à tomber à peu près perpendiculairement à la droite, et on prendra pour point de rencontre le milieu du très petit segment intercepté par l'aiguille. Je répète l'expérience un nombre de fois assez grand pour que je puisse voir si la plupart des fréquences se massent autour d'une fréquence limite.

On se trouvera en présence du même problème, si l'on recherche la probabilité qu'un accident qui a eu lieu entre Bordeaux et Paris se soit produit sur une partie déterminée du parcours, par exemple sur le kilomètre qui traverse Poitiers (1[re] probabilité), ou (2[e] probabilité) sur le kilomètre qui se trouve en pleine campagne à mi-distance de Poitiers et de Chatellerault.

Ou encore, en admettant qu'on ne puisse pas discerner sur le papier deux points distants de moins d'un $\frac{1}{10}$ de millimètre par exemple, on tire au sort la distance d'un point variable à un point fixe exprimée à $\frac{1}{10}$ de millimètre près. Par exemple, je suppose que la droite a 10 centimètres. On peut y distinguer mille dixièmes de millimètres. Donc il y a mille points distincts sur cette droite. Il s'agit d'en déterminer un par le hasard. Nous prenons 10 boules numérotées de 0 à 9, et nous faisons trois tirages, de façon à obtenir un nombre de trois chiffres (dont plusieurs pourront être des zéros). Ce nombre représentera la distance au point choisi. —

Dans chacun des exemples envisagés je peux vérifier expérimentalement quelle probabilité il y a pour un point ainsi déterminé d'être situé sur un segment déterminé AB. Même si, dans ces différents exemples, on donne à AB la même longueur, on trouvera des valeurs différentes de la probabilité.

Courbe de probabilité. — Pour bien mettre en évidence comment la valeur de la probabilité varie avec la catégorie d'épreuves envisagées, nous allons introduire les *courbes de probabilité.*

Revenons au cas général, en supposant que l'événement en question obéit aux lois du hasard, pour tout segment donné. Alors considérons trois points A B C sur la droite ; à chacun des segments AB, BC, AC correspond une probabilité déterminée, dont la valeur est comprise entre 0 et 1. On peut les représenter provisoirement par les notations $\underline{AB}$, $\underline{BC}$, $\underline{AC}$ (où l'ordre de ces points est indifférent pour chaque probabilité). Maintenant, si, par exemple, B est entre A et C, l'événement qui consiste en ce que le point mobile M tombe entre A et C peut se produire de deux manières : il tombe 1° entre A et B, 2° entre B et C. Ces deux manières sont incompatibles d'ailleurs. Par conséquent, d'après le théorème des probabilités totales, on a

$$\underline{AC} = \underline{AB} + \underline{BC}$$

Ceci permet de représenter la probabilité correspondant à un intervalle AB d'une façon plus simple. $\underline{AB}$ est un nombre qui ne dépend que de A et de B, qui est

fonction de deux points. Si je prends un point O en dehors des positions possibles de M, j'aurai

$$\underline{OB} = \underline{OA} + \underline{AB}$$

ou

$$\underline{OA} = \underline{OB} + \underline{BA}$$

suivant que A est entre O et B ou B entre O et A. Dans les deux cas

$$\underline{AB} = \underline{BA} = |\,\underline{OB} - \underline{OA}\,|\ (^1)$$

Je suppose O fixe. Je vois que $\underline{OC}$ (ou probabilité pour M d'être entre O et C) est une quantité qui est « fonction » de la position du point C. On peut la désigner par F(C). On aura donc :

$$\underline{AB} = |\,F(B) - F(A)\,|$$

Ainsi la probabilité de tomber sur AB peut être mise sous la forme de la différence arithmétique des valeurs prises par certaine fonction d'un point seulement.

Indiquons comment on pourrait interpréter plus simplement la quantité F(C).

Considérons la droite OO′, et supposons qu'on la divise en un certain nombre de parties égales c_1, c_2, ...

Imaginons qu'elle soit remplacée par une barre rectiligne, de densité variable quand on s'avance vers la

1. On désigne d'une façon générale par $|\,a - b\,|$ la différence arithmétique entre le plus grand et le plus petit des nombres a et b, par opposition à $a - b$ ou $(a - b)$ qui représente la même différence précédée du signe $+$ ou $-$ suivant que a est plus grand ou plus petit que b. Par exemple $|\,3 - 4\,| = 1$; $3 - 4 = -1$.

droite. Considérons le segment c_{k+1}. Supposons que la densité d soit égale au quotient : $\frac{Pc_k}{c_k}$, où Pc_k est la probabilité que M soit sur c_k. Le poids de c_k sera la probabilité correspondante.

Le poids de chaque segment (par exemple du segment composé de c_2, c_3, c_4, c_5) exprimera la probabilité correspondante. La somme des poids de divers segments sera égale à la somme des probabilités correspondantes Cela est vrai si A et B sont certains des points C_1, C_2, ... qui limitent les segments c_1, c_2 ..... Sinon A sera compris par exemple entre C_k et C_{k+1}, B entre C_h et C_{h+1}. On aura :

$$\underline{C_k C_{h+1}} > \underline{AB} > \underline{C_{k+1} C_h}$$

La différence entre la probabilité de AB et le poids donné — par excès ou par défaut — sera, en valeur absolue, inférieure ou égale à $\underline{C_k C_{h+1}} - \underline{C_{k+1} C_h} = \underline{C_k C_{k+1}} + \underline{C_h C_{h+1}}$.

Je fais l'hypothèse que la répartition des probabilités est telle que, si le segment est suffisamment petit, la probabilité est aussi petite qu'on voudra (tend vers zéro).

Supposons qu'on prenne un très grand nombre d'intervalles égaux. $C_k C_{k+1}$ et $C_h C_{h+1}$ seront très petits. D'après l'hypothèse précédente il en sera de même de $\underline{C_k C_{h+1}}$ et $\underline{C_h C_{h+1}}$. La différence entre $\overline{AB}$ et le poids de AB sera aussi petite qu'on voudra : le poids de AB sera égal à la probabilité que nous cherchons.

Ainsi le calcul des densités successives m'aura per

mis de définir la densité d'une barre en chacun de ses points de telle façon que le poids de la barre soit égal à la probabilité qui correspond au segment envisagé.

On peut appeler densité moyenne de la probabilité sur un segment de droite le quotient de la probabilité sur un segment par la longueur de celui-ci. Soit $f(x)$ la densité de la probabilité *au point* x, c'est-à-dire sur un segment assez petit contenant x. Construisons la courbe L qui représente graphiquement $f(x)$ et qu'on peut appeler *la courbe de densité des probabilités* ou plus brièvement la courbe de probabilité. Quelle sera la forme de cette courbe ?

Remarquons que l'aire limitée par un segment AB de Ox, par les ordonnées en A et B (c'est-à-dire les perpendiculaires à Ox en A et B) et par la courbe des probabilités, est égale à la probabilité qu'a le point M d'être sur AB. Car si on considère cette aire comme représentant le poids de AB dans une seconde répartition de la densité sur AB, la densité d'un petit segment AB de Ox serait égale à cette aire divisée par AB, ce qui est approximativement égal à $f(x)$. De sorte que les deux répartitions sont bien identiques.

1° Cas des tirages au sort. S'il y a 1.000 points sur le segment, pour chaque point on a une probabilité $\frac{1}{1.000}$. La densité sera la même partout. La courbe L sera une droite parallèle à l'axe des x (fig. 1). Le poids de AB et par conséquent $\overline{AB}$ (ou la probabilité de AB) sera proportionnel à la longueur de AB. On peut dire

que, dans ce cas, *la répartition de la probabilité sur la droite Ox est homogène.*

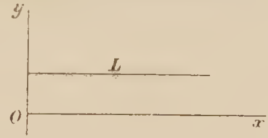


Fig. 1.

2° Dans le cas où nous cherchons la courbe de probabilité des accidents entre Bordeaux et Paris, nous remarquons que sur deux parcours égaux, l'un en rase campagne, l'autre en ville, il y a plus d'accidents dans la ville (et de même toutes les fois qu'une raison spéciale augmente les risques en un point du parcours). On aura donc une courbe très

Fig. 2.

irrégulière qui pourra affecter la forme de la figure 2, forme schématique, où on n'a pas observé les proportions.

3° Dans le cas où on jette une aiguille perpendicu-

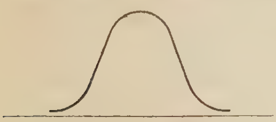


Fig. 3.

Fig. 4

lairement à une droite, si, par exemple, on s'applique à la jeter près d'un point donné, la probabilité sera plus grande pour le segment le plus proche de ce point. L'ex-

périence montre qu'on a une courbe dite en cloche, comme dans la figure 3.

Si, au contraire, on ne vise particulièrement aucun point, comme malgré tout le jeu ne peut s'étendre indéfiniment dans les deux sens, on aura une courbe comme dans la figure 4.

Très souvent, quand on parle de la probabilité pour un point d'une droite de se trouver sur un segment donné, sans qu'on spécifie la façon matérielle dont les épreuves seront exécutées, comme rien ne distingue alors un segment d'un autre, on est amené naturellement à admettre que la répartition des probabilités est homogène. Il est important de remarquer que c'est là une hypothèse, et non une convention, que cette hypothèse est toute gratuite, et qu'en fait elle se révèle généralement fausse.

On trouve cependant dans bien des cas des résultats exacts en l'acceptant, parce qu'on ne l'applique que dans la région où elle est sensiblement exacte, par exemple dans la région DE de la figure précédente. Il peut même arriver qu'on obtienne des résultats sensiblement exacts, alors même que l'hypothèse n'est réalisée nulle part, parce que, comme l'a montré Poincaré, certains problèmes de probabilités continues ont une solution indépendante de la loi de répartition des probabilités, pourvu que celle-ci satisfasse à des conditions de continuité extrêmement générales. C'est là une des découvertes capitales de Poincaré, dans le domaine des probabilités.

On le vérifierait à propos du problème, qui a été le premier exemple de problème des probabilités continues, de l'aiguille de Buffon. Mais l'analyse en est un peu difficile. Commençons par celui que Poincaré s'est proposé de résoudre.

Problème de la roulette. — On considère une roue mobile autour d'un axe, qui porte sur ses bords des divisions successivement rouges et noires (en nombre égal de chaque couleur). On lance cette roulette, et, au bout d'un certain temps, elle s'arrête. L'une des divisions se trouve en face d'un repère fixe. Quelle est la probabilité pour que cette division soit noire ?

On suppose que les divisions rouges sont égales entre elles, et les noires de même. Les rouges sont avec les noires dans un rapport constant : chaque division noire vaut k fois une division rouge.

Un premier raisonnement consistera à supposer que, si l'on a soin de ne pas donner une impulsion trop petite, les probabilités sont homogènes, que, si la roulette a tourné d'un certain angle θ, la probabilité que cet angle soit compris dans un intervalle i est proportionnel à i (c'est-à-dire à la longueur de l'arc) : en d'autres termes la probabilité pour qu'un arc vienne se placer en face du repère est la même pour tous les arcs de même longueur. La probabilité étant p pour chaque division rouge sera pk pour chaque division noire.

Pour l'ensemble des divisions rouges (en nombre n) elle sera $Q = np$.

Pour l'ensemble des divisions noires (en nombre n) elle sera $P = nkp$.

Donc $P = kQ$. On en déduit immédiatement la valeur de P et de Q.

La somme est égale à 1, car ce sont deux probabilités complémentaires. Donc $1 = Q + P = Q + kQ = (1 + k)\,Q$, d'où $Q = \dfrac{1}{1+k}$.

En particulier dans le cas où (comme le suppose Poincaré) les divisions rouges et noires sont égales, on a $k = 1$; d'où $Q = \dfrac{1}{1+1} = \dfrac{1}{2}$, $P = 1 - Q = \dfrac{1}{2}$; ainsi $P = Q$.

Plaçons-nous maintenant dans le cas le plus général,

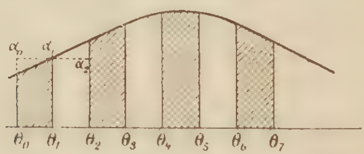


Fig. 5.

où la courbe de probabilité n'est pas connue avec précision. Soit θ l'angle dont a tourné la roue au moment où elle s'est arrêtée, et soit $\delta(\theta)$ la densité correspondante de la probabilité. Reportons-nous à la courbe des probabilités (fig. 5). Marquons de hachures les régions R_1, R_2, etc., correspondant aux divisions rouges. Appelons N_1, N_2, etc., les régions correspondant aux divisions noires.

La probabilité Q que l'arrêt se produise devant une

rouge sera la somme des aires hachurées. P sera la somme des autres.

Calculons le rapport de ces aires, par exemple de R_1 et de N_1. La partie principale de R_1, c'est le rectangle $\theta_0\,\theta_1\,\alpha_1\,\alpha_0$.

La partie principale de N_1, c'est le rectangle $\theta_1\,\theta_2\,\alpha_2\,\alpha_1$. Ces rectangles sont de même hauteur h, et si la base du premier est b, celle de l'autre est par hypothèse kb. Les aires de ces rectangles sont donc hb et khb. L'expression $N_1 - kR_1$, si on y remplace N_1 et R_1 par leurs valeurs approchées khb et hb, deviendra $khb - k.\,hb = 0$. De même pour $N_2 - kR_2$.

Donc on peut remplacer approximativement N_1 par kR_1, etc., et $\frac{P}{Q}$ par $\frac{kR_1 + kR_2 + \ldots\ldots}{R_1 + R_2 + \ldots\ldots} = k$.

On voit que k est une valeur approchée de $\frac{P}{Q}$, et d'autant plus approchée que les divisions sont plus petites. Et ce résultat est indépendant de la courbe de probabilité. En effet, je ne me suis point servi des probabilités homogènes, je n'ai pas supposé que la courbe des probabilités était une droite.

Paradoxe de Bertrand. — L'exemple suivant, dû à Bertrand, montre bien qu'il est nécessaire dans les problèmes de probabilités continues de spécifier ce qu'on entend par « choisi au hasard ».

Cherchons la probabilité pour qu'une corde AB, tracée au hasard dans un cercle, ait une longueur plus petite que le côté du triangle équilatéral inscrit. — Bertrand,

voulant montrer que les résultats sont douteux, indique trois solutions et trois résultats numériquement différents [1].

On peut dire : « Si l'une des extrémités de la corde est connue, ce renseignement ne change pas la probabilité ; la symétrie du cercle ne permet d'y attacher aucune influence, favorable ou défavorable à l'arrivée de l'événement demandé. L'une des extrémités de la corde étant connue, la direction doit être réglée par le hasard. Si l'on trace les deux côtés du triangle équilatéral ayant pour sommet le point donné, ils forment entre eux et avec la tangente trois angles de 60°. La corde, pour être plus grande que le côté du triangle équilatéral, doit se trouver dans celui des trois angles qui est compris entre les deux autres. La probabilité pour que le hasard entre trois angles égaux qui peuvent le recevoir le dirige dans celui-là, semble par définition égale à $\frac{1}{3}$.

« On peut dire aussi : si l'on connaît la direction de la corde, ce renseignement ne change pas la probabilité. La symétrie du cercle ne permet pas ..... La direction de la corde étant donnée, elle doit, pour être plus grande que le côté du triangle équilatéral, couper l'un ou l'autre des rayons qui composent le diamètre perpendiculaire, dans la moitié la plus voisine du centre. La probabilité pour qu'il en soit ainsi semble par définition, égale à $\frac{1}{2}$.

« On peut dire encore : choisir une corde au hasard

1. Bertrand, *Calcul des Probabilités*, p. 4 et 5.

c'est en choisir au hasard le point milieu. Pour que la corde soit plus grande que le côté du triangle équilatéral, il faut et il suffit que le point milieu soit à une distance du centre plus petite que la moitié du rayon, c'est-à-dire à l'intérieur d'un cercle quatre fois plus petit en surface. Le nombre des points situés dans l'intérieur d'une surface quatre fois moindre est quatre fois moindre. La probabilité pour que la corde dont le milieu est choisi au hasard soit plus grande que le côté du triangle équilatéral semble, par définition, égale à $\frac{1}{4}$.

« Entre ces trois réponses, quelle est la véritable? Aucune des trois n'est fausse, aucune n'est exacte, la question est mal posée ».

L'incertitude dans laquelle Bertrand laisse le lecteur, Poincaré l'a dissipée en disant : « nous avons fait des hypothèses différentes » et en montrant que la réponse dépend de la nature d'une certaine fonction que nous ignorons et qui reste arbitraire. Cette fonction est celle qui joue dans le problème actuel le rôle de la densité de la probabilité dans le cas d'un point mobile sur une droite. Poincaré ajoute : il faut nous *donner* cette fonction « au début du problème par une convention spéciale pour qu'il ait un sens ». En d'autres termes, le problème n'a de sens que si nous convenons que la répartition des probabilités est connue et donnée. Et cela est parfaitement juste.

Mais il faut bien remarquer que si l'on veut *appliquer* la solution dans un cas concret déterminé, ce n'est plus

une *convention* qu'il y aura lieu de faire, ce sera une *expérience*. Il faudra mesurer la répartition des probabilités dans le cas concret envisagé ; on pourra alors voir si cette répartition est conforme à l'une des trois répartitions considérées successivement par Bertrand, ou à aucune d'elles. Suivant le cas, la réponse sera l'un des trois nombres obtenus, ou aucun d'entre eux. La situation est ici exactement la même, dans un cas plus général, que lorsque nous avons voulu calculer la probabilité, pour un point d'une droite, de se trouver sur un segment donné. Nous avons donné des exemples qui montrent que même dans ce cas simple la question est « mal posée », qu'il faut, pour la résoudre, « convenir » d'une certaine loi de variation de la densité des probabilités, mais que, pour appliquer la solution, il fallait non plus une convention (ni une hypothèse) mais une connaissance expérimentale précise de la répartition des probabilités dans le cas concret envisagé.

Pour en revenir au cas de Bertrand, on voit bien maintenant que le paradoxe consiste en ce que, suivant la réalisation matérielle du choix de la corde, la répartition des probabilités ne sera plus la même, et par conséquent la solution sera différente.

Par exemple, supposons que les extrémités de la corde soient déterminées en lançant une aiguille comme à la roulette, mais deux fois de suite. Alors la première solution de Bertrand conviendra et la probabilité cherchée sera $\frac{1}{3}$.

Calcul des probabilités dépendant de la position d'un point dans un plan (ou de l'ensemble de deux nombres, de deux variables). — Si je jette une pièce au hasard et si je prends le centre de la pièce, la position de ce centre sur le sol est déterminée par deux nombres : par exemple ses distances à deux axes perpendiculaires tracés sur le sol. Si je trouve au hasard les deux extrémités d'une corde en lançant deux fois de suite une aiguille qui s'arrête en face d'une division d'un cadran, la situation de la corde est déterminée par deux nombres : les deux angles décrits par l'aiguille. — Ainsi, si un événement est déterminé par la positiond'un point ou d'une droite dans un plan, on peut aussi le considérer comme déterminé par deux nombres. — Pour étudier la probabilité d'un événement qui dépend de deux variables, nous suivrons la méthode indiquée pour les probabilités linéaires (position d'un point sur une droite, déterminée par un seul nombre). Soit le point M qui a une certaine position dans le plan T. Quelle est la probabilité pour que M soit dans un contour déterminé S ?

Si on précise la façon dont le point M est déterminé, il faut chercher expérimentalement la fréquence de cet événement. Mais si nous avons à traiter ce problème successivement pour plusieurs contours σ, il vaut mieux chercher une méthode générale (expérimentale), et en tirer une règle mathématique pour trouver la probabilité pour chaque contour σ.

Nous divisons le plan par un carrelage serré. Le point M sera assujetti à rester dans un certain nombre

de ces carrés (par exemple, si on jette une pièce dans une chambre fermée, les carrés extérieurs ne sont pas accessibles). Nous admettrons qu'à chacun d'eux correspond une probabilité (que les fréquences y tendent vers une limite, ce qui est une hypothèse). Appelons σ_1 un des carrés, et p_{σ_1} la probabilité correspondante. Nous appellerons *densité moyenne* de la probabilité dans ce carré le quotient : $\frac{p_{\sigma_1}}{\sigma_1}$.

Supposons que, sur chaque carré, on construise un prisme qui ait comme hauteur la densité moyenne $d_{\sigma_1}, \ldots$ On obtient ainsi un certain volume, dont nous appellerons la surface supérieure S.

Soit maintenant σ_i l'ensemble des petits carrés intérieurs à σ et σ_c l'ensemble de ceux de ces carrés qui ne sont pas extérieurs à σ (fig. 6). Soit p_{σ_i}, p_{σ_c}, p_σ les probabilités respectives pour que le point appartienne à σ_i, σ_c ou σ. On a évidemment

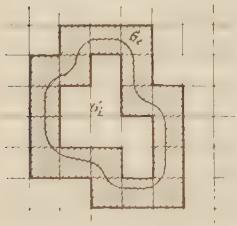


Fig. 6.

$$p_{\sigma_i} \leqq p_\sigma \leqq p_{\sigma_c}$$

Or soient V_{σ_i}, V_{σ_c}, V_σ les volumes limités d'une part par le plan T et par la surface S, d'autre part par les cylindres perpendiculaires au plan T et ayant pour bases respectives σ_i, σ, σ_c. On a évidemment :

$$V_{\sigma_i} \leqq V_\sigma \leqq V_{\sigma_c}$$

Mais, d'après le théorème des probabilités totales, p_{σ_i}, p_{σ_c} sont les sommes des probabilités relatives à chacun des petits carrés de σ_i et de σ_c. Et la hauteur de chaque petit prisme, $d_{\sigma_1}, \ldots$ a été choisie de sorte que le volume de ce petit prisme $\sigma_1 d_{\sigma_1}, \ldots$ soit mesuré par la probabilité correspondante $p_{\sigma_1}, \ldots$

Donc :

$$p_{\sigma_i} = V_{\sigma_i}; \ p_{\sigma_c} = V_{\sigma_c}$$

Donc :

$$V_{\sigma_i} \leqq p_\sigma \leqq V_{\sigma_c}$$

Donc :

$$| p_\sigma - V_\sigma | \leqq V_{\sigma_c} - V_{\sigma_i} = p_{\sigma_c} - p_{\sigma_i} = P_{\sigma'}.$$

Car $p_{\sigma_c} - p_{\sigma_i}$, c'est la probabilité $P_{\sigma'}$, pour que le point soit dans l'aire qu'on peut appeler $\sigma' = \sigma_c - \sigma_i$, aire formée par l'ensemble des petits carrés qui ont un point commun avec le contour de σ. Si σ est une courbe suffisamment régulière, la superficie $\sigma' = \sigma_c - \sigma_i$ (couverte de hachures sur la figure) tend vers zéro quand le carrelage est plus serré. Donc $P_{\sigma'}$ tend aussi vers zéro[1]. On aura à la limite p_σ = limite de V_σ.

On peut donc construire une certaine surface (non plane en général), Σ, limite de S, qui jouit de la propriété suivante : quelle que soit l'aire σ du plan T,

1. Notre raisonnement (et ses conséquences) suppose donc que la probabilité $P_{\sigma'}$, pour que le point M tombe dans σ' est très petite si l'*aire* σ' est très petite. Cette hypothèse n'est pas aussi naturelle qu'elle paraît au premier abord, et les cas ne sont pas rares où elle n'est pas admissible. Nous n'aborderons pas ici l'examen plus difficile de ces cas.

la probabilité correspondante p_σ est égale au volume limité par le plan T, la surface Σ, et le cylindre projetant σ.[1]

Si la répartition des probabilités est homogène dans le plan, Σ sera un plan parallèle à T. La probabilité p_σ sera proportionnelle à l'aire σ.

Problème du franc carreau. — On jette un disque sur un carrelage plan illimité. Supposons que les dalles soient des carrés tous égaux. Si la pièce avait un diamètre D plus grand que le côté a du carreau A, quel que soit l'endroit où tombe la pièce, elle serait à cheval sur plus d'un carreau. Si au contraire D est plus petit que A, elle pourra être aussi à l'intérieur de l'un des carreaux. Trouver dans ce cas la probabilité p pour qu'elle tombe à l'intérieur d'un carré quelconque.

Portons notre attention sur la situation du centre de la pièce, C. Il faut que les distances du point C aux quatre côtés du carré soient supérieures à $\frac{D}{2}$. Dessinons à l'intérieur du carré (fig. 6 *bis*) des parallèles à ses côtés à une distance $\frac{D}{2}$. C doit tomber dans le carré intérieur en pointillé. Il en est de même pour tout autre carreau.

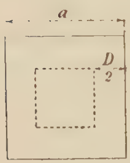


Fig. 6 *bis*.

Si on jette la pièce à la main, les points où peut tomber la pièce sont dans une région limitée. Soit N

1. Σ est la surface de probabilité qui représente la répartition de la densité de la probabilité dans le plan comme la courbe de probabilité représente la répartition correspondante sur une droite.

le nombre de carreaux supposé fini sur lesquels il est possible que la pièce tombe. La pièce peut tomber dans une surface Σ d'aire Na^2.

Appelons Σ' la somme des aires des petits carrés pointillés correspondant à chacun de ces carreaux. Supposons la répartition des probabilités homogène. Si P_Σ est la probabilité pour le centre C de se trouver dans Σ, et $P_{\Sigma'}$, la probabilité de se trouver dans Σ', on a

$$\frac{P_{\Sigma'}}{P_\Sigma} = \frac{\Sigma'}{\Sigma}$$

$$\Sigma = Na^2, \quad \Sigma' = N\left(a - 2\frac{D}{2}\right)^2 = N(a-D)^2 \ .$$

Donc

$$\frac{\Sigma'}{\Sigma} = \frac{N\ (a-D)^2}{N\ a^2} = \frac{(a-D)^2}{a^2} \ .$$

Mais P_Σ, c'est la probabilité pour qu'il tombe sur le carrelage : c'est la certitude $= 1$. D'autre part $P_{\Sigma'} = p$.

Donc

$$\frac{P_{\Sigma'}}{P_\Sigma} = p,$$

Ainsi

$$p = \left(\frac{a-D}{a}\right)^2 .$$

Par exemple, pour que la pièce tombe aussi souvent à l'intérieur d'un carreau qu'à cheval sur le carrelage, dans un grand nombre de jets, il faut que $p = \frac{1}{2}$; d'où $\frac{1}{\sqrt{2}} = \frac{a-D}{a} = 1 - \frac{D}{a}$; $\frac{D}{a} = 1 - \frac{1}{\sqrt{2}} = 0{,}3$;

le diamètre de la pièce doit être égal aux $\frac{3}{10}$ du côté des carreaux.

Critique. — Ici, comme précédemment, notre hypothèse n'est pas exacte, si la position du centre de la pièce au lieu d'être attribuée à un hasard idéal est déterminée par un jet. Dans ce cas il est évident qu'il y aura une dégradation insensible dans la fréquence par centimètre carré quand on s'approche de la région où la pièce ne tombe jamais. Il ne peut y avoir homogénéité absolue. Toutefois on peut, par un artifice, montrer que le résultat sera très voisin du précédent si la probabilité, n'étant pas homogène, satisfait à des conditions assez simples de continuité.

Supposons que les carreaux soient assez petits pour que la densité de la probabilité par unité de surface ne varie pas beaucoup dans chacun d'eux, sauf près du bord de la région accessible où elle décroît rapidement vers zéro. Appelons p_1, p_2,.... les probabilités pour que le centre de la pièce tombe dans les carreaux numérotés 1, 2,.... et p'_1, p'_2,.... pour qu'elle tombe dans les carrés fictifs intérieurs correspondant à chacun d'eux. On aura alors :

$$\frac{p'_1}{p_1} = \left(\frac{a - D}{a}\right)^2 ; \quad \frac{p'_2}{p_2} = \left(\frac{a - D}{a}\right)^2 ; \ldots$$

Si on a soin de numéroter les carrés en commençant par le centre et en terminant par ceux qui sont situés au loin, l'égalité précédente sera surtout exacte pour les

premiers, mais d'autre part p_N, $p_{N-1}, \ldots$ seront très petits. Or

$$p = p'_1 + p'_2 + \ldots \quad \text{et} \quad 1 = p_1 + p_2 + \ldots$$

Les premières quantités p'_1, $p'_2, \ldots\ldots$ sont respectivement très voisines de $p_1 \left(\frac{a-D}{a}\right)^2$, $p_2 \left(\frac{a-D}{a}\right)^2, \ldots\ldots$ Les quantités p'_n, $p'_{n+1}, \ldots\ldots$ p'_N quand n est grand ne sont pas nécessairement dans un rapport avec $p_n, \ldots\ldots$ p_N qui soit voisin de $\left(\frac{a-D}{a}\right)^2$, mais d'autre part, comme elles sont très petites, on ne fait pas d'erreur absolue sensible en les remplaçant par

$$p_1 \left(\frac{a-D}{a}\right)^2 \ldots p_n \left(\frac{a-D}{a}\right)^2, \ldots$$

Donc on a, très approximativement

$$p = p_1 \left(\frac{a-D}{a}\right)^2 + p_2 \left(\frac{a-D}{a}\right)^2 + \ldots$$
$$= \left(\frac{a-D}{a}\right)^2 \left[p_1 + p_2 + \ldots\right] = \left(\frac{a-D}{a}\right)^2.$$

Nous voyons que cette fois, en supposant que les carreaux soient assez petits par rapport à la région accessible pour qu'il y ait à peu près homogénéité sauf aux bords, on n'a pas non plus une égalité rigoureuse, mais une égalité seulement approchée.

En faisant 50 fois l'expérience avec une pièce de 37 millimètres de diamètre, sur des carreaux de 165 millimètres de côté, nous avons trouvé que la pièce est tombée 29 fois à l'intérieur de l'un des carreaux ; la

fréquence de l'événement favorable a été $\frac{29}{50} = 0,58$. La valeur calculée théoriquement par notre formule est $\left(\frac{165 - 37}{165}\right)^2 = 0,60$. La différence pourrait être plus grande sans qu'il y ait à s'en étonner, étant donné le petit nombre d'épreuves.

EXERCICES[1]

I (F). — Reprendre le problème du franc carreau en supposant les carreaux alternativement noirs et blancs, et calculer la probabilité pour que la pièce tombe à l'intérieur d'un carreau blanc.

II (F). — Etant donnée la répartition des successions en 1907 indiquée page 11, tracer approximativement la courbe de densité de la probabilité pour que l'actif d'une succession prise au hasard soit compris entre deux nombres déterminés. On fera une première figure en supposant la densité constante dans chacun des intervalles de valeur de l'actif x considérés dans le tableau numérique. Et pour calculer la densité moyenne de la probabilité dans un intervalle, on admettra que les fréquences sont des valeurs suffisamment approchées des probabilités.

Ayant ainsi tracé une courbe discontinue formée de segments de droite, on admettra que la courbe cherchée est une courbe continue qui donne approximativement les mêmes densités moyennes que la courbe discontinue.

La courbe obtenue décroît très rapidement et le dessin en est très approximatif.

III (F). — Question analogue à la précédente, mais relative

1. Voir la note (1) de la page 11.

au tableau de la page 53. Dessiner la courbe de densité de la probabilité pour qu'une commune prise au hasard ait eu, en 1921, une population comprise entre deux nombres donnés. Pour préciser et faciliter la construction nous compléterons le tableau de la page 53, en ajoutant que 42, 900, 8 et 4 des communes recensées avaient respectivement de 1 à 100, de 101 à 500; de 10.001 à 20.000, plus de 20.000 habitants.

CHAPITRE III

PROBABILITÉ DES HYPOTHÈSES (OU DES CAUSES)

1er *exemple. Le problème des trois coffrets* (indiqué par Bertrand). — Trois coffrets O, M, A, d'apparence identique contiennent chacun deux tiroirs. Dans O, le premier tiroir contient une pièce d'or, et le second, une pièce d'or ; dans M, le premier contient une pièce d'or, le second, une pièce d'argent ; dans A, le premier, une pièce d'argent, le second, une pièce d'argent. On tire au hasard un tiroir, et on y trouve une pièce d'or. Quelle est la probabilité pour que l'autre tiroir du même coffret contienne aussi une pièce d'or ?

Quand on a trouvé la pièce d'or, deux hypothèses étaient possibles : c'est le coffret O ou le coffret M.

Quelle probabilité y a-t-il pour que ce soit O ? Il n'y avait que deux coffrets contenant au moins une pièce d'or. La probabilité est la même, semble-t-il, pour l'un et l'autre, soit $\frac{1}{2}$. Ce raisonnement est incorrect, car on a choisi non pas seulement le coffret, mais dans le coffret le tiroir. On pouvait tirer un tiroir quelconque.

En réalité, l'événement qu'est l'ouverture d'un tiroir contenant une pièce d'or étant survenu, la catégorie d'épreuves dans laquelle il s'agit de chercher la probabilité x comporte trois cas également possibles. D'autre part il y a deux cas favorables (premier tiroir de O, deuxième tiroir de O). Donc $x = \frac{2}{3}$. Ce raisonnement est correct. Mais il n'est pas général. Mettons-le sous une forme telle qu'on puisse l'appliquer au problème général qui va nous occuper de la probabilité des causes.

Soit y la probabilité d'avoir ouvert un tiroir de O. Cette probabilité est celle de l'événement qui consiste dans le concours des deux événements suivants : choix d'un tiroir où se trouve une pièce d'or ; choix d'un tiroir appartenant à O. Appliquons le théorème des probabilités composées en considérant successivement chacun de ces événements comme le premier. La probabilité y est donc égale au produit de la probabilité du premier, multipliée par la probabilité du second dans la catégorie d'épreuves où le premier a eu lieu ; donc $y = \frac{3}{6} \times x$. Mais on peut renverser l'ordre des événements ; la probabilité y est alors égale au produit de la probabilité du second (choix d'un tiroir de O) par celle du premier (choix d'un tiroir où se trouve une pièce d'or quand on a choisi un tiroir de O), — soit la certitude ; donc

$$y = \frac{1}{3} \times 1 = \frac{1}{3}.$$

On a, en égalant les deux valeurs de y :

$$\frac{3}{6}x = \frac{1}{3} \qquad 3x = 2 \qquad x = \frac{2}{3}.$$

2^e^ *exemple, tiré de la statistique.* — Dans une certaine population, plusieurs personnes meurent pendant l'année. On demande la probabilité P pour qu'un des morts pris au hasard ait l'âge de 35 ans.

Ceci suppose un certain nombre d'hypothèses ou de données, dont la première est que la probabilité de mourir pendant l'année, pour cette population, est déterminée.

Soit q_x le taux annuel de mortalité à l'âge x dans cette population, c'est-à-dire la probabilité pour une personne d'âge x de mourir pendant l'année. Dans cette population, il y a N personnes, d'âges divers. Soit N_x le nombre de celles des N personnes qui sont d'âge x. $\frac{N_x}{N}$ est la fréquence de l'âge x de ce groupe. Si N est nombreux, $\frac{N_x}{N}$ est pratiquement la probabilité pour une personne du groupe d'avoir l'âge x, soit π_x.

Appliquons le raisonnement général précédent. P est la probabilité cherchée, pour qu'un des morts pris au hasard dans le groupe ait eu 35 ans au début de l'année. L'événement survenu E consiste en ce que cette personne est morte dans l'année. J'introduis la probabilité auxiliaire y : que l'événement ait lieu sous la condition que la personne soit âgée de 35 ans.

$y =$ probabilité pour une personne du groupe d'avoir

35 ans $\times$ probabilité pour une personne de 35 ans de mourir dans l'année $= \pi_{35} \times q_{35}$.

Renversons l'ordre :

$y =$ probabilité pour une personne du groupe de mourir dans l'année $\times$ probabilité pour un mort du groupe d'avoir 35 ans $= \mathrm{T} \times \mathrm{P}$.

Calculons T. Une personne du groupe peut avoir l'âge 0, 1, 2, ... x, ... et les probabilités de mourir dans l'année à ces différents âges (qui s'excluent mutuellement) sont les quantités q_0, q_1, ... q_x, ... La probabilité pour une de ces personnes d'avoir l'âge x est la quantité π_x. Donc (voir page 34),

$\mathrm{T} = \Sigma\,(q_x \pi_x)$ ou la somme des quantités $q_x\,\pi_x$, soit $\pi_0 q_0 + \pi_1 q_1 + \ldots\ldots$

Finalement

$$q_{35} \times \pi_{35} = y = \mathrm{P} \times \Sigma\,(q_x\,\pi_x). \tag{13}$$

Donc

$$\mathrm{P} = \frac{q_{35}\,\pi_{35}}{\Sigma\,(q_x\,\pi_x)} = \frac{q_{35}\,\pi_{35}}{q_0\,\pi_0 + q_1\,\pi_1 + \ldots}$$

Mais quelles valeurs faut-il prendre pour les quantités π_0, π_1, ... ? On a ici un exemple de l'indétermination apparente qui se présente dans les problèmes de probabilité des hypothèses.

Dans le cas actuel je pourrais être tenté de supposer que tous les âges sont également possibles, puisque je ne sais rien sur ces quantités : les probabilités π_0, π_1, ... seraient donc égales. La formule (13) donnerait P

$\dfrac{q_{35}\pi_0}{\pi_0 q_0 + \pi_0 q_1 + \cdots} = \dfrac{q_{35}}{q_0 + q_1 + \cdots}$. Mais ce raisonnement est insuffisant. Il faut en effet spécifier *la catégorie des épreuves* dans lesquelles la probabilité doit être calculée. Supposons qu'on constitue le groupe de toute la population de la France. Il y aura beaucoup plus d'enfants au-dessous d'un an que de personnes de 70 ans. La répartition entre les différents âges est publiée par l'*Annuaire de la Statistique générale de la France*. Pour en donner une idée, il suffira d'indiquer les chiffres suivants pour le recensement de 1911. — La proportion des Français du sexe masculin était, pour cent, de 1,91 de 0 à 1 ans ; 32,73 de 1 à 19 ans ; 30,82 de 20 à 39 ans ; 22,97 de 40 à 59 ans ; 10,81 de 60 à 79 ans ; 0,76 au-dessus de 79 ans. La valeur de π_x va donc en décroissant de 0,0191 pour $x = 0$, à $\dfrac{0,3273}{19}$ $= 0,0172$ pour une certaine valeur de x entre 1 et 19 ; $\dfrac{0,3082}{20} = 0,0154$ pour une certaine valeur de x entre 20 et 39 ;..... $\dfrac{0,1081}{20} = 0,0054$, entre 60 et 79. On voit que ces nombres sont très différents, et quelle erreur on commettrait en les supposant égaux.

Problème général. — On considère un événement fortuit E, qu'on examine dans une catégorie d'épreuves déterminée. Supposons qu'il puisse se présenter sous les modalités E_1, E_2,..... E_N, seules possibles et incompatibles entre elles. Supposons que chaque modalité E_k soit réalisée par le concours de l'événement E et d'une circonstance H_k. Par exemple E sera la mort d'un individu

dans une certaine année, H_k le fait pour cet individu d'avoir au début de l'année un âge déterminé, E_k l'ensemble de ces deux hypothèses.

Ceci posé, l'événement E étant survenu, on demande quelle est la probabilité P_1 qu'il soit survenu sous l'hypothèse H_1, par exemple, connaissant les probabilités p_1, p_2,.... p_n pour que E se produise quand chacune des hypothèses H_1, H_2.... est respectivement réalisée, et les probabilités π_1, π_2, π_3,.... π_n pour que ce soient les circonstances H_1, H_2,.... H_n qui se soient produites.

J'introduis une probabilité auxiliaire, la probabilité y_1 pour qu'à la fois E et H_1 aient lieu, et j'applique le théorème des probabilités composées de deux façons (EH_1, puis H_1E).

On peut d'abord considérer y_1 comme égal à la probabilité pour que E ait lieu $\times$ la probabilité pour que E, ayant eu lieu, ait eu lieu sous l'hypothèse H_1. La seconde quantité est P. La première est la somme des probabilités pour que E ait lieu sous les diverses hypothèses.

Donc

$$y_1 = (p_1 \pi_1 + p_2 \pi_2 + \ldots).\ P.$$

Prenons l'ordre inverse. y_1 est la probabilité de l'hypothèse $H_1 \times$ la probabilité de E sous l'hypothèse H_1 $= \pi_1\ p_1$. En égalant les deux valeurs de y_1, on a :

$$p_1 \times \pi_1 = (\pi_1 p_1 + \pi_2 p_2 + \ldots). \times P.$$

D'où la *formule de Bayes* (appliquée surtout par Laplace).

$$P = \frac{p_1 \pi_1}{p_1\pi_1 + p_2\pi_2 + \dots} \tag{14}$$

Précautions à prendre dans l'emploi de la formule de Bayes. — Pour donner une idée des difficultés d'application auxquelles peut donner lieu la formule de Bayes, lorsque les probabilités π des diverses hypothèses avant l'événement sont insuffisamment connues, considérons l'exemple suivant, traité par Poincaré.

« A l'écarté, votre adversaire donne et retourne le roi ; quelle est la probabilité pour que ce soit un Grec ? » — L'événement E, c'est l'acte de retourner le roi. Les deux hypothèses sont : que votre adversaire soit un honnête homme : hypothèse H_1, ou qu'il soit un Grec : hypothèse H_2. On demande la probabilité de H_2, E ayant eu lieu.

P_1, ou la probabilité pour un honnête homme de retourner un roi d'un jeu de 32 cartes, est $\frac{4}{32} = \frac{1}{8}$; P_2, ou ce que devient P_1 pour un Grec, sera évidemment plus grand que P_1 : mais on sait qu'il est compris entre $\frac{1}{8}$ et 1. Si le moment est décisif, s'il y a un fort enjeu, on peut supposer avec Poincaré que le Grec s'arrangera pour retourner la plus forte carte ; il y a certitude : $P_2 = 1$. S'il s'agit au contraire d'un coup quelconque, comme le Grec n'a pas avantage à éveiller

l'attention par des gains trop fréquents, P_2 ne sera pas beaucoup plus proche de 1 que de $\frac{1}{8}$; on pourra prendre par exemple, avec M. Borel, $P_2 = \frac{1}{4}$: on voit que c'est affaire d'appréciation. Reste à évaluer π_1 et π_2 ; π_1 et π_2 sont les fréquences respectives des honnêtes gens et des Grecs parmi les personnes que je rencontre au jeu : j'estime que ces Grecs sont rares, π_2 est petit, et π_1 est très voisin de 1. — On a alors pour la probabilité cherchée :

$$P = \frac{P_2 \pi_2}{P_1 \pi_1 + P_2 \pi_2} = \frac{P_2 \pi_2}{\frac{1}{8} \pi_1 + P_2 \pi_2}.$$

Ici P_2 est compris entre 1 et $\frac{1}{8}$; π_2 est très petit ; donc $P_2 \pi_2$ l'est aussi ; par suite $\frac{1}{8} \pi_1 + P_2 \pi_2$ ne diffère pas beaucoup de $\frac{1}{8} \pi_1$; on peut prendre approximativement pour P, puisque π_1 est voisin de 1, la valeur :

$$\frac{P_2 \pi_2}{\frac{1}{8}} = 8 P_2 \pi_2 > 8 \frac{1}{8} \pi_2 = \pi_2$$

puisque $P_2 > \frac{1}{8}$. On constate, comme il est naturel, que la probabilité pour que mon adversaire soit un Grec s'est accrue quand il a retourné la plus forte carte. Si l'on se trouve dans le cas de Poincaré, $P_2 = 1$, $P = 8\pi_2$, la probabilité est 8 fois plus forte. Dans le cas de

Borel, $P_2 = \frac{1}{4}$, $P = 2\pi_2$, la probabilité est seulement deux fois plus forte.

Ce résultat semble paradoxal : après tout, mon adversaire n'a point changé du fait qu'il a joué. Mais distinguons les catégories d'épreuves. Supposons qu'un médium puisse nous dire après chaque jeu si nous avons joué avec un Grec ou non. Notons alors toutes les parties que nous avons jouées dans une longue période avec un grand nombre de personnes, et relevons la proportion de Grecs dans ce nombre de personnes : ce sera environ π_2. Maintenant, notons le nombre de ces parties qui se sont terminées l'adversaire retournant un roi, et relevons la proportion de Grecs dans ce nombre (plus petit) de parties : ce sera environ P. Il ne paraît plus du tout paradoxal que la proportion de Grecs dans le second cas soit plus forte que dans le premier, que $P > \pi_2$: c'est un résultat de bon sens. — Mais il résulte de la formule de Bayes que la proportion, dans le second cas, ne sera pas 8 fois plus grande que dans le premier. Plus exactement, si la même série d'expériences est faite par un certain nombre de personnes, dans la plupart de ces séries on trouvera que le rapport des fréquences est < 8. Non seulement le calcul ne conduit pas à un paradoxe, non seulement il est conforme au bon sens, mais il fournit une limite précise que le seul bon sens serait impuissant à prévoir.

EXERCICE

(F) Dans une première génération la probabilité pour une mère de famille d'avoir 1, 2, 3 ..... enfants prend les valeurs respectives :

$\pi_1 = 0{,}053$; $\pi_2 = 0{,}057$; $\pi_3 = 0{,}1$; $\pi_4 = 0{,}132$; $\pi_5 = 0{,}140$; $\pi_6 = 0{,}124$; $\pi_7 = 0{,}113$; $\pi_8 = 0{,}092$; $\pi_9 = 0{,}076$; $\pi_{10} = 0{,}052$; $\pi'_{11} = 0{,}061$

où π'_{11} est la probabilité d'avoir plus de 10 enfants.

Dans une seconde génération la probabilité pour une mère de famille d'avoir quatre enfants est :

$$p_1 = \frac{5}{53}, \quad p_2 = \frac{5}{57}, \quad p_3 = 0{,}19, \quad p_4 = \frac{17}{132},$$
$$p_5 = \frac{21}{140}, \quad p_6 = \frac{15}{124}, \quad p_7 = \frac{18}{113}, \quad p_8 = \frac{10}{92},$$
$$p_9 = \frac{14}{76}, \quad p_{10} = \frac{1}{52}, \quad p'_{11} = \frac{7}{61}$$

suivant qu'elle appartenait elle-même à une famille de 1,2, ... 10 ou plus de 10 enfants.

On demande de calculer la probabilité P_3 pour que dans la seconde génération une mère de quatre enfants ait appartenu à une famille de trois enfants.

Nota. — On a pris ici comme valeurs des probabilités π et p les fréquences correspondantes du tableau de la page 123. Ce tableau permettra donc de vérifier l'exactitude de la valeur à calculer pour P_3.

CHAPITRE IV

L'ESPÉRANCE MATHÉMATIQUE

Supposons qu'un certain nombre de personnes désirent acheter des tableaux. Elles réunissent une somme suffisante, et les tirent au sort. Admettons qu'un tableau vaille 5.000 francs. Il y a 1.000 personnes.

Avant le tirage, chacune a la même espérance de gagner, et paie une somme x. Donc

$$1.000\, x = 5.000 \text{ francs}, \quad x = \frac{5.000}{1.000} = 5 \text{ francs}.$$

Ici il n'y a pas de doute : le chiffre de la cotisation est évident.

Considérons un autre groupe de personnes qui désirent mettre en réserve une certaine somme pour le cas d'accident, soit 5.000 francs par accident. Chacune verse une cotisation. Quelle sera-t-elle ? Cela dépend du nombre des accidents, qui variera suivant que le groupe est plus ou moins nombreux. — S'il est peu nombreux, pendant plusieurs années de suite il n'y aura aucun accident, et, une année, il y en aura un, deux ou trois, qui pourront coûter cher à chacune des

personnes du groupe. S'il y a 10 personnes seulement, elles auront à payer 0 franc les premières années, et, la première année où il y aura un accident, $\frac{5.000}{10}$, soit 500 francs. Si le nombre des personnes est grand, $=$ M, si m est le nombre des accidents dans l'année, l'expérience montre que la fréquence $\frac{m}{M} = F$ varie assez peu. La somme à payer par personne sera :

$$\frac{m \times 5.000}{M} = 5.000 \times F$$

soit une somme sensiblement constante, puisque F a une valeur à peu près fixe et voisine de la probabilité de l'événement. Il en résulte qu'au lieu de se trouver complètement dans l'incertitude, les M personnes sauront d'avance que la cotisation à verser sera approximativement égale au paiement aléatoire $\times$ la probabilité de l'événement qui détermine ce paiement. Ce résultat est général. Il s'applique par exemple aussi bien dans le cas du tirage au sort du tableau. La probabilité de gagner est $\frac{1}{1.000}$; la somme à payer par personne sera en effet : $5.000 \times \frac{1}{1.000}$.

Je suppose qu'un événement E est considéré comme un événement favorable qui détermine, lorsqu'il est réalisé, le paiement d'une somme S. Il s'agit de savoir, lorsqu'un grand nombre de personnes M pourraient se voir attribuer ce paiement, combien chacune devra verser. Soit m le nombre des personnes pour qui l'évé-

nement E s'est réalisé, x la cotisation, Mx, la somme des cotisations, $= m$S, ou le total des sommes payées. D'où

$$x = \frac{m\mathrm{S}}{\mathrm{M}} = \mathrm{S}\,\frac{m}{\mathrm{M}} = \mathrm{SF}.$$

Si le nombre des épreuves est très grand et pris au hasard dans une catégorie d'épreuves où la probabilité de l'événement est P, F sera voisin de P. Donc x sera voisin de PS.

Si nous considérons l'état d'esprit de ces personnes avant l'épreuve, nous trouvons que chacune a les mêmes raisons, puisque la probabilité est la même pour toutes, d'espérer qu'elle touchera la somme S. Si elle veut vendre sa chance d'obtenir S à une autre personne, celle-ci devra naturellement lui payer sa cotisation. La mesure de cette espérance est égale à la valeur de la cotisation à verser pour s'assurer S avec la probabilité P. D'où ce théorème (qui n'est pas une définition, mais qui exprime la relation entre les notions de probabilité et d'espérance mathématique) :

L'espérance mathématique d'un gain éventuel est égale au produit de ce gain par la probabilité de le réaliser.

Généralisation. — Il peut se présenter des cas où il y a plusieurs sommes dont le paiement dépend de la réalisation de plusieurs événements. Si E_1 se produit, on convient de payer S_1, si E_2 se produit, S_2, ces événe-

ments ayant les probabilités p_1, p_2, etc. Les versements sont indépendants les uns des autres, les événements peuvent ne pas l'être. Il s'agit de chiffrer l'espérance mathématique d'une personne dans cette hypothèse.

S'il faut fixer une cotisation x_1 pour assurer le paiement de S_1, x_2 pour S_2, etc., la cotisation à verser pour assurer l'ensemble $S_1 + S_2 + \ldots$ sera $= x_1 + x_2 + \ldots$ *L'espérance mathématique correspondant au paiement des diverses sommes S_1, S_2, ..., est la somme des espérances mathématiques correspondant au paiement de chacune de ces sommes $= S_1p_1 + S_2p_2 + \ldots$, chacun de ces paiements devant être considéré isolément.* — Par exemple, un joueur, Pierre, s'engage à payer à Paul 2 francs s'il amène un point pair au jeu de dés, et *en outre* 1 franc s'il amène 6. S'il amène 6, il lui paiera 1 franc + 2 francs; s'il amène 5, il ne paiera rien; 4, il paiera 2 francs; 3, 0 franc; 2, 2 francs; 1, 0 franc. Quelle est l'espérance mathématique de Pierre? La probabilité d'amener 6 étant $\frac{1}{6}$, et d'amener un nombre pair, $\frac{3}{6}$ ou $\frac{1}{2}$, l'espérance mathématique sera :

$$1 \times \frac{1}{6} + 2 \times \frac{1}{2} = \frac{7}{6} \text{ d'un franc.}$$

Problème de l'aiguille. — Appliquons cette notion d'espérance mathématique à la résolution d'un problème déja signalé, le problème de l'aiguille, proposé par Buffon. — On jette une aiguille sur un plan sur lequel

sont tracés des traits parallèles et équidistants. Soit a la distance qui sépare les traits, l la longueur de l'aiguille. Si $l < a$, l'aiguille ne pourra croiser qu'un des traits (si elle en croise un). Plaçons-nous dans ce cas, où une seule intersection est possible. Calculons la probabilité p pour que l'aiguille jetée au hasard soit à cheval sur un des traits quel qu'il soit.

On peut résoudre ce problème par le calcul intégral. Mais une solution ingénieuse, due à Barbier, permet de s'en passer. Elle suppose essentiellement une répartition homogène des probabilités. — Supposons qu'on considère l'aiguille comme divisée en morceaux, chacun d'une longueur λ. Si je jette chacun de ces morceaux, la probabilité sera la même pour tous. Appelons E l'espérance mathématique d'un joueur à qui on promet un franc par intersection. — l est quelconque, $<$ ou $> a$. Dans le cas où $l < a$, comme il n'y a qu'une intersection possible, il y a à chaque épreuve 0 ou 1 intersection, par suite $\mathrm{E} = 1 \times p = p$. Si $l > a$, il peut y avoir plus d'une intersection, et on a $\mathrm{E} > p$. — Pour chaque petit morceau λ, il y a une espérance mathématique $= \varepsilon$, et si $l = m\lambda$, on a $\mathrm{E} = m\varepsilon$.

Donc $\dfrac{\mathrm{E}}{\varepsilon} = \dfrac{l}{\lambda}$, d'où $\dfrac{\mathrm{E}}{l} = \dfrac{\varepsilon}{\lambda}$. Mais nous remarquons que le même raisonnement s'applique, que l'aiguille soit droite ou polygonale. Si donc on a une aiguille polygonale composée de m' petits morceaux λ, si l' est sa longueur totale, E' l'espérance mathématique correspondante, on aura $\dfrac{\mathrm{E}'}{l'} = \dfrac{\varepsilon}{\lambda} = \dfrac{\mathrm{E}}{l}$. Or l'égalité $\dfrac{\mathrm{E}}{l} = \dfrac{\mathrm{E}'}{l'}$ est indépen-

dante du nombre de morceaux λ ; elle s'applique donc, à la limite, à une aiguille courbe.

Calculons alors l'espérance E' pour une aiguille ayant la forme d'une circonférence de diamètre a.

Cette circonférence rencontrera toujours une des parallèles en deux points. Donc $E' = 2$, puisqu'à chaque fois le joueur gagne 2 francs. — Or la longueur de cette aiguille circulaire sera $l' = \pi a$. Donc :

$$\frac{E}{l} = \frac{E'}{\pi a} = \frac{2}{\pi a}, \quad E = \frac{2l}{\pi a}.$$

Si $l < a$, on a $E = p$, donc $p = \frac{2l}{\pi a}$. Nous obtenons la probabilité cherchée.

Dans la solution précédente, l'hypothèse sur la répartition homogène des probabilités n'est pas entièrement justifiée. En s'inspirant de la solution qu'a donnée Poincaré du problème de la roulette, M. Hostinsky a montré que, pour une répartition très générale des probabilités satisfaisant seulement à des conditions simples de continuité, la valeur de p est, non pas nécessairement égale à $\frac{2l}{\pi a}$, mais, pour une valeur donnée de $\frac{l}{a}$, d'autant plus voisine de ce rapport que les parallèles sont plus rapprochées, ou encore que leur distance est plus petite par rapport à la région accessible à l'aiguille.

Dans ces conditions on voit qu'on ne doit pas s'attendre à trouver par l'expérience une fréquence f égale

à $\frac{2l}{\pi a}$, puisque non seulement f n'est que voisin de p, mais aussi p n'est que voisin de $\frac{2l}{\pi a}$. En admettant cependant l'égalité approximative[1] : $f \triangleq \frac{2l}{\pi a}$, on en déduit $\pi \triangleq \frac{2l}{fa}$, et par suite on a un moyen assez curieux de calculer expérimentalement le nombre π par une suite d'épreuves où ne figure aucun cercle. On mesure la distance a des parallèles, la longueur l de l'aiguille, on compte sur N jets le nombre m d'épreuves avec intersection, et on a $f = \frac{m}{N}$, $\pi \triangleq \frac{2l}{fa}$.

L'astronome suisse Wolf a effectué 5.000 épreuves avec une aiguille l ayant 36 millimètres, et des traits distants de 45 millimètres. Il a obtenu : $\pi \triangleq 3.1596$. — L'Italien Lazzerini a fait 2.000 épreuves ($a = 30$ millimètres, $l = 25$ millimètres), et a obtenu : $\pi \triangleq 3,1446$. — Un autre Italien, Reina, a fait 2.520 épreuves ($a = 60$ millimètres, $l = 30$ millimètres) ; il a obtenu : $\pi \triangleq 2,93$, et, en tenant compte de la largeur de l'aiguille, il a corrigé cette valeur, et obtenu : $\pi \triangleq 2,90$. Ces résultats sont assez peu différents de la valeur exacte de π : $\pi = 3,14159.....$ — On remarquera que les résultats les meilleurs, conformément au raisonnement de M. Hostinsky, correspondent à la plus petite valeur de a.

1. On a souvent avantage à employer la notation A $\triangleq$ B pour indiquer que A et B diffèrent peu l'un de l'autre.

Applications de l'espérance mathématique aux problèmes d'assurance. — Avant de chercher à évaluer l'espérance mathématique d'un paiement incertain, évaluons la valeur actuelle d'un paiement futur certain, compte tenu des intérêts composés.

Paiement certain. — Si une somme de 1.000 francs est placée à intérêts composés au taux de 5 p. 100, elle s'augmente de $1.000 \times \frac{5}{100}$ au bout d'un an ; elle est donc devenue $1.000 + 1.000 \times \frac{5}{100} = 1.000 \left(1 + \frac{5}{100}\right) = 1.000 \times 1,05$. Au bout de la seconde année, elle est devenue $1000\ (1,05)^2$. Au bout de 3, 4,..... 10 années, elle sera devenue $1.000\ (1,05)^3$; $1.000 \times (1,05)^4$;..... $1.000\ (1,05)^{10}$. — D'une façon générale appelons i l'intérêt produit par 1 franc, c'est-à-dire que si T est le taux p. 100, $i = \frac{T}{100}$; alors 1 franc devient $1 + i$, $(1 + i)^2$..... au bout de 1, 2..... années. Donc une somme V deviendra $S = V\ (1 + i)^n$ au bout de n années. — Inversement quelle est la somme V qu'il faut verser actuellement pour payer S au bout de n années au taux i ? Ce sera une somme liée à S par la relation :

$$V = \frac{S}{(1 + i)^n} ;$$

V est la valeur actuelle d'un capital S différé de n années.

Assurances. — Supposons qu'un certain nombre de personnes s'adressent à une Compagnie pour contracter une assurance à capital différé. Moyennant le paiement d'une prime, au bout de n années, l'assuré, s'il est en vie, recevra une somme S. Calculons la prime due par l'assuré : il s'agit de la prime *pure*, c'est-à-dire de la prime nécessaire pour que la Compagnie se procure cette somme, — à quoi il faudra ajouter un surplus pour couvrir les frais, pertes, etc. Nous supposerons en outre que l'assuré désire se libérer en une fois, c'est-à-dire qu'il paie une prime unique au moment de la signature du contrat.

Si la société n'avait que peu d'assurés, il lui serait difficile d'indiquer la valeur de la prime. Si, au contraire, elle a un grand nombre d'assurés, elle a pu établir des tables indiquant avec une approximation suffisante pour chaque âge x le nombre l_x de survivants dans un groupe donné. Alors s'il y a l_x personnes désirant chacune un capital différé au bout de n années, et si chacune paie V francs, la société aura reçu $l_x \times$ V francs, et au bout de n années elle aura versé $l_{x+n} \times$ S. Mais dans l'intervalle elle aura aussi touché les intérêts des primes, de sorte que la somme reçue par la société est devenue : $l_x \mathrm{V} (1 + i)^n$, qui doivent lui permettre de verser $\mathrm{S}l_{x+n}$.

D'où $\mathrm{V} = \frac{l_{x+n}}{l} \cdot \frac{\mathrm{S}}{(1+i)^n}$. Or $\frac{\mathrm{S}}{(1+i)^n}$ est la valeur actuelle du paiement aléatoire, $\frac{l_{x+n}}{l_x}$ est la probabilité de ce paiement. Donc la prime pure est l'espérance mathé-

matique de recevoir la valeur actuelle du capital différé.

Dans la pratique on conduit ainsi les calculs. On écrit : $V = vS$.

v étant la prime pure, à l'âge x, pour un capital de 1 franc différé de n années ; v peut s'écrire :

$$v = \frac{l_{x+n}}{l_x} \cdot \frac{1}{(1+i)^n} = \frac{l_{x+n}}{(1+i)^{x+n}} : \frac{l_x}{(1+i)^x} = \frac{D_{x+n}}{D_x}.$$

Et, pour chaque taux i on dresse des tables dites de commutation de $D_x = \dfrac{l_x}{(1+i)^x}$.

On n'a plus alors qu'à chercher dans les tables D_x, D_{x+n}, d'où v, et ensuite $V = vS$.

On n'a donc à consulter qu'une table (de commutation) au lieu de deux (d'intérêts et de survie).

Si l'assuré veut obtenir *une rente viagère au bout de* n *années*, on considère cette rente viagère de montant annuel R comme une suite d'assurances différées d'un montant constant R, payable au bout de n, $n + 1$,..... années. La valeur de la prime pure unique sera donc, pour un assuré d'âge x :

$$\frac{R D_{x+n}}{D_x} + \frac{R D_{x+n+1}}{D_x} + \ldots$$
$$= R \left[\frac{D_{x+n} + D_{x+n+1} + \ldots}{D_x} \right]$$

La suite s'arrête d'elle-même puisque les tables de mortalité n'indiquent plus de survivants au delà de 100 à 110 ans. On peut alors dresser d'avance des

tables donnant les valeurs de sommes N_y telles que $D_y + D_{y+1} + \ldots\ldots$ pour chaque valeur de y.

Ces indications sur les assurances ne sont données qu'à titre d'exemple.

EXERCICES

I (F). — Calculer les nombres de commutation D_{30}, D_{40} qui correspondent aux âges de 30 et 40 ans pour un taux d'intérêt de 5 p. 100 l'an, lorsque la table de survie employée est la table dite R. F. qui donne

$$l_{30} = 771.681, \qquad l_{40} = 717.338.$$

En déduire la prime unique pure à payer à 30 ans pour un capital de 10.000 francs différé de 10 ans en cas de vie.

II. — Trouver la formule qui donne la prime annuelle pure P_x à payer pendant la vie entière à partir de l'âge x en remplacement d'une prime unique U_x à payer à l'âge x.

CHAPITRE V

LA NOTION D'ÉCART ET LES VALEURS TYPIQUES D'UN ENSEMBLE DE NOMBRES

1re Section

Cas d'un nombre fini d'épreuves

On peut dire que, si on considère un ensemble de paiements aléatoires, l'espérance mathématique caractérise par un nombre précis l'ensemble de ces paiements; plus exactement, elle précise une de leurs caractéristiques; elle répond à la condition suivante : c'est la valeur qu'un entrepreneur (de jeux) devrait payer, sans gain ni perte, à chacun des joueurs pour acheter son espérance de gagner une certaine somme, pour s'assurer l'ensemble de tous les paiements qui devront être faits à tous les joueurs.

Nous nous proposons maintenant un problème plus général : nous voulons caractériser, dans la mesure du possible, un ensemble de nombres, y_1, y_2, ... y_n. se rapportant à un même objet, à un même ordre de données, par un nombre unique, que nous appellerons *valeur typique* de cet ensemble.

Supposons d'abord que ces nombres sont déterminés par un certain ensemble d'épreuves : par exemple les nombres des points obtenus successivement pendant une ou plusieurs parties ; ou les nombres des jours de pluie dans chacun des mois de l'année. Il s'agit, au moyen d'un nombre choisi arbitrairement comme valeur typique en tenant compte des circonstances, de donner une idée de l'ordre de grandeur de l'ensemble de ces nombres. Nous admettrons d'abord, dans tout l'examen de cette question, que nous n'avons affaire qu'à *un nombre fini* N d'épreuves, et nous arriverons ainsi à certaines notions *qui seront indépendantes de la notion de probabilité*. Nous serons ainsi préparés à aborder dans la seconde section le cas d'un nombre infini d'épreuves qui suppose, pour se prêter à la généralisation, l'existence de probabilités définies.

Pour caractériser un ensemble de nombres $y_1, y_2, \ldots,$ un bon moyen semble consister à choisir un nombre m qui s'écarte le moins possible de chacun de ces nombres.

La convention la plus naturelle (mais non la meilleure) pour définir m peut être que $|y_1 - \omega|$, $|y_2 - \omega|, \ldots$ $|y_N - \omega|$ en valeurs absolues[1] soient les plus petits possible quand $\omega = m$. Mais si on rend petit un de ces écarts, les autres ne le seront peut-être

1. Rappelons encore une fois que la valeur *absolue* d'une différence $a - b$ est la différence arithmétique obtenue en retranchant la plus petite quantité de la plus grande. On la désigne par $|a - b|$, de sorte qu'on a $|4 - 3| = |3 - 4|$ alors que $4 - 3 = 1$ et $3 - 4 = -1$.

pas. D'où la notion de l'*écart d'un ensemble de nombres y avec ω*. On choisira alors pour valeur typique m la valeur de ω pour laquelle cet écart est minimum. Reste à définir cet écart.

1° *Écart maximum.* — Je pourrais appeler par exemple écart de ω et de l'ensemble des y la plus grande des différences ci-dessus indiquées, et chercher à ce qu'elle soit le plus petite possible. A quel choix cela conduirait-il pour m ?

En supposant les y rangés par rang de grandeur, ($y_1 \leqq y_2 \leqq y_3 \ldots\ldots \leqq y_N$), il est évident que l'écart maximum de ω avec les y est celui des deux nombres $| y_1 - \omega |$, $| y_N - \omega |$ qui est le plus grand. Si l'on représente ces quantités comme les distances de deux points Y_1, Y_N d'une droite à un autre point W de la même droite, où l'on appelle $y_1, y_2, \ldots\ldots y_N, \omega$ les distances des points correspondants à l'origine, on voit que l'écart de ω avec les y sera minimum quand W sera au milieu de $Y_1 Y_N$, c'est-à-dire quand ω sera égal à $\frac{y_1 + y_N}{2}$. Ainsi $\frac{y_1 + y}{2}$ est en ce sens une valeur typique de l'ensemble des y.

Mais ce choix n'est pas très satisfaisant, car, dans la plupart des cas qui se présentent en statistique, la plus petite et la plus grande des valeurs y sont des valeurs extrêmes, anormales, et il n'y a pas de raison pour admettre que les cas vraiment anormaux se présentent en même nombre, et avec le même degré d'intensité, dans le sens des valeurs plus grandes et des plus petites,

surtout dans un ensemble d'épreuves en nombre fini. Ces valeurs sont en outre celles qui seront le moins déterminées par la nature de l'ensemble des nombres $y_1, \ldots\ldots y_N$, et qui dépendront le plus de circonstances accessoires.

Considérons cependant, par exemple, *la répartition mensuelle des jours de pluie et de la quantité d'eau tombée par mois dans une année.*

Année	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Sep.	Oct.	Nov.	Déc.
	Nombre de jours de pluie.											
1917	14	9	$\overline{21}$	16	14	14	15	20	$\underline{8}$	20	14	12
1918	18	11	12	20	11	$\underline{6}$	17	12	18	12	14	$\overline{24}$
1919	20	16	24	15	$\underline{6}$	7	16	9	9	14	22	$\overline{25}$
	Hauteur de pluie en mm.											
1917	26,3	26,0	70,5	60,4	31,6	$\overline{97,3}$	72,4	52,3	26,8	75,4	33,4	$\underline{25,1}$
1918	41,8	$\underline{17,3}$	53,3	52,0	67,5	$\underline{17,3}$	54,1	35,0	$\overline{84,1}$	23,5	70,4	60,2
1919	73,9	89,2	66	65,8	17,7	$\underline{10,6}$	67,1	27,6	27,6	36,1	$\overline{92,8}$	79,5

Nous avons marqué pour chaque année le nombre maximum par un trait supérieur ($\overline{21}$ pour la première ligne) et le nombre minimum par un trait inférieur ($\underline{8}$ pour la première ligne).

Supposons que nous cherchions si une année a été plus particulièrement pluvieuse.

Ces nombres annuels, très irréguliers, donnent l'impression du plus complet désordre.

La méthode précédente appliquée ici conduit à considérer comme valeurs typiques des nombres de jours de pluie dans un mois :

$$\text{en } 1917 : \frac{8+21}{2} = 14\ \frac{1}{2} ; \text{ en } 1918 : \frac{6+24}{2} = 15 ;$$

$$\text{en } 1919 : \frac{6+25}{2} = 15\ \frac{1}{2} ;$$

la quantité de pluie en un mois :

$$\text{en } 1917 : \frac{25{,}1+97{,}3}{2} = 61^{\text{mm}}{,}2 ; \text{ en } 1918 : \frac{17{,}3+84{,}1}{2} = 50^{\text{mm}}{,}7 ;$$

$$\text{en } 1919 : \frac{10{,}6+92{,}8}{2} = 51^{\text{mm}}{,}7 .$$

On voit déjà qu'une certaine constance d'une année à l'autre se manifeste dans ces valeurs typiques, constance qui n'apparaissait nullement dans les tableaux.

2° *Ecart moyen*. — On pourrait arriver d'une autre manière au choix d'une valeur typique, et, par exemple, écrire que le nombre ω doit être choisi de façon que la somme $| y_1 - \omega | + \ldots\ldots + | y_N - \omega |$ soit le plus petite possible. Alors on considérerait l'écart entre ω et l'ensemble des y comme le plus petit possible. Mais afin d'avoir un nombre qui se réduise à l'écart $| \omega - y_1 |$ quand $y_1 , \ldots\ldots y_N$ sont égaux, on appelle

généralement écart moyen des y et de ω la somme des quantités précédentes divisée par le nombre N des y, soit :

$$\frac{|y_1 - \omega| + \ldots + |y_N - \omega|}{N}.$$

En d'autres termes, *l'écart moyen du nombre aléatoire* y *avec* ω *est la moyenne arithmétique des valeurs absolues des écarts avec* ω *des valeurs prises dans les différentes épreuves par ce nombre aléatoire.* Naturellement, si les y sont égaux, la valeur de l'écart sera bien cette fois égale à l'écart d'un quelconque des y et de ω. On choisira comme valeur typique le nombre ω pour lequel cet écart est minimum.

Dans l'ensemble des calculs, comme N est fixe, je chercherai la valeur minimum du numérateur, c'est-à-dire de la somme $S = |\omega - y_1| + \ldots.. + |\omega - y_N|$, quand ω varie Considérons deux valeurs particulières de ω, ω' et ω'' $(\omega' \leq \omega'')$.

Plaçons-nous dans le cas où, entre les deux, il n'y a aucun nombre y. D'autre part je suppose qu'on a ordonné les nombres y par valeurs croissantes :

$$y_1 \leq y_2 \ldots \leq y_k \leq \omega' \leq \omega'' \leq y_{k-1} \leq y_{k-2} \leq \ldots$$

La première somme

$$S' = (\omega' - y_1) + \ldots + (\omega' - y_k) + (y_{k+1} - \omega') + \ldots$$

La deuxième somme

$$S'' = (\omega'' - y_1) + \ldots + (\omega'' - y_k) + (y_{k+1} - \omega'') + \ldots$$

La différence entre ces deux quantités est :

$$S' - S'' = (w' - w'')\,k + (w'' - w')\,k',$$

Comme w'' est plus grand que w', cela donne :

$$S' - S'' = (w'' - w')\,(k' - k),$$

k' est le nombre des termes y après w', w''.

Par conséquent S' sera plus grand ou plus petit que S'', suivant que k' sera plus grand ou plus petit que k, $S' = S''$ si $k' = k$.

Il y a trois hypothèses : 1° il y a un plus grand nombre de termes après qu'avant ; 2° il y en a un plus petit ; 3° il y en a autant. Distinguons ces cas suivant que le nombre N des épreuves est impair ou pair. Supposons le impair, $= 2\,h + 1$.

Il y a alors un terme exactement au milieu, de rang médian $h + 1$. La première hypothèse revient alors à : w', w'' sont situés avant le terme de rang médian, et la deuxième à : w', w'' sont situés après le terme de rang médian (la troisième hypothèse ne se présente pas).

Si E est l'écart de w avec les y, quand w croît avant le terme de rang médian, l'écart E décroît ; en effet, puisque $w'' > w'$, la somme S'' est plus petite que la somme S', tant que k', ou le nombre des termes après w', w'', est plus grand que k, ou le nombre des termes avant. Quand w croît après le terme médian, l'écart E croît. Par conséquent, quand on fait croître w, l'écart commence à décroître, puis croît, et est minimum quand w est égal au terme de rang médian.

Supposons le nombre N des épreuves pair : $N = 2h$. Il y a alors dans la suite des y deux termes y_h y_{h+1} qui se placent exactement au milieu de la suite par leurs rangs, aux extrémités de ce qu'on peut appeler l'intervalle médian. S'il arrive que ces deux termes, quoique de rangs distincts, soient égaux, on est ramené au cas précédent. Sinon $k < k'$ quand $\omega < y_h$; $k > k'$ quand $\omega > y_{h+1}$, et $k = k'$ quand ω est dans l'intervalle y_h, y_{h+1}. Donc l'écart décroît d'abord, reste constant quand ω est dans cet intervalle, et croît ensuite.

Nous pouvons appeler *valeur médiane* de l'ensemble des nombres y la valeur de y qui est de rang médian, ou bien les valeurs situées dans l'intervalle de rang médian. L'*écart moyen de ω avec les y a son minimum, lorsque ω est égal à une valeur médiane de l'ensemble des y.*

L'intérêt de ce théorème, c'est qu'il dispense d'étudier d'abord la variation de l'écart avec ω quand on veut chercher son minimum pour un groupe de valeurs numériques déterminées de y. Il suffit de ranger les y par ordre de valeurs non décroissantes, puis de prendre pour valeur médiane des y soit le terme de rang médian, si le nombre N des y est impair, soit une quelconque des valeurs comprises dans l'intervalle de rang médian, quand N est pair.

Par exemple, rangeons par ordre de valeurs non décroissantes les nombres de jours de pluie par mois :

En 1917	En 1918	En 1919	Les rangs sont :
8	6	6	1
9	11	7	2
12	11	9	3
14	12	9	4
14	12	14	5
14	12	15	6
14	14	16	7
15	17	16	8
16	18	20	9
20	18	22	10
20	20	24	11
21	24	25	12

En 1917, 14 est l'unique valeur médiane ; en 1918, toute valeur entre 12 et 14, et en 1919, toute valeur entre 15 et 16 peut être prise comme valeur médiane.

L'écart minimum dans le premier cas sera :

$$\frac{6+5+2+1+2+6+6+7}{12}=\frac{35}{12}=2{,}91$$

Dans le deuxième cas, calculé à partir de 12, il sera :

$$\frac{6+1+1+2+5+6+6+8+12}{12}=\frac{47}{12}=3{,}91$$

Dans ce même deuxième cas, calculé à partir de 14, il sera :

$$\frac{8+3+3+2+2+2+3+4+4+6+10}{12}=\frac{47}{12}=3{,}91,$$

le même que nous connaissions d'avance.

Calculé à partir de toute autre valeur entre 12 et 14, par exemple de 13, ce sera : 3,91 encore.

Dans le troisième cas, ce sera : 5,25.

On peut donc dire, si l'on veut, qu'il est tombé de la pluie dans chaque mois pendant des nombres de jours dont la valeur médiane est la même pour les années 1917 et 1918, = 14, mais qu'en gros il y a eu une plus grande irrégularité en 1918 qu'en 1917, puisque l'écart moyen à partir de la valeur médiane (ou écart moyen minimum) a été d'environ 4 jours (3 jours 91 en 1918), au lieu de 3 jours (2 jours 91 en 1917).

Supposons que la suite des valeurs possibles des y soit discontinue, ce qui a lieu par exemple lorsqu'on cherche la valeur typique du nombre de maisons possédées par un propriétaire, ou du nombre d'enfants par famille, dans une ville ou dans un pays. Si on se trouve alors dans le cas exceptionnel indiqué où il y a un intervalle médian, si, par exemple il y a exactement (cas exceptionnel) le même nombre de propriétaires possédant une maison que de propriétaires possédant deux maisons ou plus ; s'il y a encore (cas non moins exceptionnel) exactement autant de familles avec un ou deux enfants que de familles avec plus de deux enfants, dans la région où se placerait la médiane, la valeur médiane est indéterminée. Cet inconvénient disparaît quand il s'agit d'une valeur qui varie d'une manière continue. Par exemple si l'on examine la statistique de la taille des conscrits on trouvera pour tout nombre de milli-

mètres voisin de la valeur médiane de la taille au moins quelques conscrits dont la taille aura ce nombre de millimètres. L'incertitude sera donc de 1 millimètre, c'est-à-dire ne sera pas plus grande que celle qui résulte des conditions de rapidité où s'effectuent les mensurations.

Calcul de la médiane. — Les quantités y_1, y_2 ..... y_N ne sont pas nécessairement différentes. Elles prennent un certain nombre $n \leqq N$ de valeurs distinctes x_1, x_2, ... x_n, chacune pouvant être prise une ou plusieurs fois, x_1 étant pris r_1 fois, x_2, r_2 fois, ... La fréquence de la valeur x_1 dans N épreuves est $\frac{r_1}{N} = f_1$; de x_2, $\frac{r_2}{N} = f_2$, ... Pour trouver l'écart moyen d'une quantité ω avec les y, nous prenons la moyenne arithmétique des quantités $| x_h - \omega |$ répétées autant de fois que x_h figure dans la catégorie d'épreuves, c'est-à-dire r_h fois. L'écart moyen de ω avec les y est donc :

$$\frac{r_1 \mid x_1 - \omega \mid + \ldots + r_n \mid x_n - \omega \mid}{N} = f_1 \mid x_1 - \omega \mid + f_2 \mid x_2 - \omega \mid + \ldots + f_n \mid x_n - \omega \mid ,$$

si l'on appelle en général f_h la fréquence $\frac{r_h}{N}$ de la valeur x_h parmi les N épreuves qui fournissent y_1, y_2, ..,.. y_N.

Cherchons maintenant la valeur médiane ; $r_1 + r_2 + \ldots + r_n$ est la somme des N nombres des épreuves.

Envisageons deux cas : si N est impair, le terme médian sera tel qu'il y ait autant de termes avant et après, dans la suite de y. Soit y_s le terme médian, s'il est égal

à x_h, le nombre p des termes placés avant y_s est la somme $r_1 + r_2 + \ldots\ldots + r_{h-1}$, augmentée peut être d'un certain nombre de fois que x_h se trouve répété avant y_s. De même le nombre des termes placés après y_s est $r_{h+1} + \ldots\ldots r_n$, plus un nombre de termes égal au nombre de fois que x_h se trouve répété après y_s. Donc $r_1 + \ldots\ldots + r_{h-1} \leqq r_h + \ldots\ldots + r_n$. Au contraire $r_1 + \ldots\ldots + r_h \geqq r_{h+1} + \ldots\ldots + r_n$.

Par suite $r_1 + \ldots\ldots + r_{h-1} \leqq \frac{N}{2}$; $r_1 + \ldots\ldots r_h \geqq \frac{N}{2}$.

Si on fait successivement les sommes r_1, $r_1 + r_2$, ... on trouvera une somme $r_1 + \ldots\ldots + r_{h-1}$, qui sera inférieure ou égale à $\frac{N}{2}$, et une autre, $r_1 + \ldots\ldots + r_h$, qui sera supérieure à $\frac{N}{2}$. On s'arrêtera à ce moment. D'où un moyen pratique de trouver la valeur médiane, qui consiste à dépasser juste la moitié du nombre des épreuves, en supposant les valeurs rangées par ordre de grandeur.

Exemple. — Cherchons la valeur médiane des durées de vie, valeur dite encore *vie probable*. Soit une table de mortalité l_0, l_1, ... l_{x-1}, l_x, l_{x+1}, ..., qui nous donne le nombre des survivants l_x à l'âge x d'un groupe déterminé. Les valeurs appelées précédemment x_0, x_1, ... sont données ici par les durées de vie. La valeur 0 année (moins d'un an) se présente dans $l_0 - l_1$ cas, la valeur 1 année, dans $l_1 - l_2$ cas, etc.

Soit r_0, r_1, ... les nombres des cas. — Calculons r_0, $r_0 + r_1$, ..., c'est-à-dire $l_0 - l_1$, $l_0 - l_2$, etc. On

pourra s'arrêter au terme tel qu'en passant au suivant on dépasse la moitié du nombre des épreuves. Il est plus simple de partir de l'âge le plus élevé ω, en les rangeant par ordre de grandeurs décroissantes : ω, $\omega - 1$, $\omega - 2$, etc., ou par exemple 100, 99, 98, etc.

On aura $r_{100} = l_{100} = 0$; $r_{99} = l_{99} - l_{100}$; $r_{98} = l_{98} - l_{99}$; ..... D'où :
$r_{100} = 0$; $r_{100} + r_{99} = l_{99}$; $r_{100} + r_{99} + r_{98} = l_{98}$; $r_{100} + r_{99} + r_{98} + r_{97} = l_{97}$; ..... Il suffira que je prenne l'indice x de l assez grand pour avoir une somme l_x qui atteigne la première $\frac{l_0}{2}$.

Soit à étudier la mortalité en France en 1901. Le nombre des membres du groupe qu'on va suivre à partir de moins de 1 an, soit $l_0 = 100.000$. Le terme médian l_x sera voisin de 50.000. Je trouve $l_{56} = 50.802$; $l_{57} = 49.682$. Le 50.000^e^ et le 50.001^e^ des membres du groupe rangés suivant leur durée de vie croissante ont 56 ans accomplis et moins de 57. 56 ans sera la valeur médiane, parce que je représente les durées de vie par un nombre exact d'années.

On peut évaluer approximativement le nombre de jours, en plus de 56 ans, qu'ils ont vécu. Il est mort dans l'année 1.120 personnes de 56 ans. Admettons que, dans ce court intervalle d'une année, il en soit mort à peu près autant chaque jour, soit $\frac{1\,020}{365} = 2{,}79$. Or, pour arriver au 50.000^e^, il faut ajouter, aux 49.198 personnes mortes à 55 ans au plus, $50.000 - 49.198 = 802$. Comme il en meurt 2,79 par jour, il suffira

d'ajouter à 56 ans un nombre de jours égal à $\frac{802}{2.79} =$ 288, pour avoir l'âge approximatif de la 50.000^e personne.

Ou, à un mois près, 56 ans 9 mois par défaut. Ce sera la vie dite probable.

Méthode graphique. — On peut procéder d'une façon plus précise par la méthode graphique. Pour obtenir par exemple la taille médiane de 100.000 conscrits, on décrit une courbe qui représente la variation du nombre S_x des conscrits ayant une taille inférieure à un nombre x de centimètres donné, par exemple pour $x = 154$, 162, 163, ... soit 9 valeurs. On a donc une courbe qui passe par 9 points connus. Ceci permet de la tracer assez bien. On détermine ensuite graphiquement la valeur de x pour le point de la courbe qui correspond à $S_x = 50.000$; ce point correspond à une valeur de x qui est la taille médiane cherchée.

Déviation. — On conçoit bien que, suivant que l'écart moyen à partir d'un nombre w est plus ou moins petit, les nombres y seront plus ou moins resserrés autour de ce nombre w. On peut préciser cette impression de la façon suivante.

Considérons une variable y, et appelons E son écart moyen à partir d'une valeur w. Je vais chercher la fréquence des valeurs y dont l'écart avec w surpasse une quantité donnée ε, soit f_ε.

Je substitue à f_ε la quantité suivante : F_t, ou fréquence

des y qui diffèrent en valeur absolue de w d'au moins t fois l'écart moyen des y avec w, ou de tE. En prenant $\varepsilon = t$E, on aura $f_\varepsilon = F_t$. On a

$$NE = |y_1 - w| + \ldots + |y_N - w|.$$

Parmi ces écarts, il y en a qui sont supérieurs à tE (d'autres non). Je suppose qu'il y en ait r et j'appelle $y'_1, y'_2, \ldots\ldots y'_r$; les valeurs d'y correspondantes dans $|y'_1 - w|, |y'_2 - w| \ldots\ldots |y'_r - w|$. On a $|y'_1 - w| + |y'_2 - w| + \ldots\ldots + |y'_r - w| \geqq rt$E ; à plus forte raison aura-t-on :

$$NE \geqq rtE \qquad \text{ou} \qquad N \geqq rt. \text{ Donc } \frac{r}{N} \leqq \frac{1}{t}.$$

Mais $\frac{r}{N}$ est la fréquence F_t des y' ; donc *la fréquence, dans un groupe de nombres* y, *de ceux de ces nombres dont les écarts individuels avec un nombre* w *sont supérieurs ou égaux à un multiple donné* t *de l'écart moyen de ces* y *avec* w *est inférieure ou égale à l'inverse de ce multiple* t.

Par exemple considérons la hauteur de pluie tombée par mois en 1917 (page 102). La médiane est 52 millimètres. L'écart moyen E $= 21^{mm},5$. La fréquence des cas où la hauteur de pluie tombée par mois est supérieure ou inférieure à 52 millimètres d'au moins une fois et demi $21^{mm},5$ sera $F_{1,5} \leqq \frac{1}{1,5}$ ou $\frac{2}{3}$.

Autrement dit, il y aura, au plus, deux tiers des mois (huit mois) où l'écart sera supérieur à une fois et demi $21^{mm},5$.

Assurément, on pourrait déterminer directement le nombre de mois cherché en le comptant sur le tableau. Mais si on a un tableau étendu, la règle précédente permettra d'éviter un comptage long et minutieux.

3° *Écart quadratique moyen.* — On peut donner de l'écart une définition un peu différente, moins naturelle, mais qui conduit à une définition de la valeur typique plus naturelle, et à laquelle on s'est généralement arrêté, avec une préférence souvent excessive. Soit les valeurs y_1, y_2, ... y_N. On forme la somme des carrés des différences :

$$(y_1 - \omega)^2 + \ldots + (y_N - \omega)^2.$$

Pour la nouvelle valeur typique de l'ensemble des y, cette somme devra être le plus petite possible. On pourrait très bien convenir de mesurer la grandeur de l'écart entre ω et cet ensemble de valeurs y_1, y_2, ... y_N par la valeur de cette somme. Mais il y a avantage à chercher une quantité qui, lorsque les y seront tous égaux, sera égale à $(y - \omega)$ en valeur absolue. C'est pourquoi on convient généralement d'appeler écart quadratique moyen de ω avec l'ensemble des y la quantité :

$$\sqrt{\frac{(y_1 - \omega)^2 + \ldots + (y_N - \omega)^2}{N}}.$$

Pour quelle valeur de ω cet écart est-il minimum ? Avant d'en venir à la formule qui répond à cette ques-

tion, il convient d'énoncer une remarque très importante, due à Bienaymé, retrouvée plus tard et utilisée systématiquement par Tchebicheff (qui n'a pas su que Bienaymé l'avait indiquée). (Ces deux auteurs se sont d'ailleurs bornés au cas d'un nombre infini d'épreuves que nous traiterons plus loin.)

Lemme de Bienaymé. — Etant donnée une valeur ω, et un ensemble de valeurs y, chercher la fréquence des valeurs y qui diffèrent de ω d'au moins ε (ε étant une quantité donnée) soit f_ε. Je me sers de la valeur de l'écart quadratique moyen à partir de ω, et je transforme le problème : connaissant l'écart quadratique moyen E des y à partir de ω, chercher la fréquence F_t des valeurs de y qui diffèrent de ω d'au moins tE.

J'appelle y'_1, ..., y'_n, celles des valeurs de y qui diffèrent de ω d'au moins t E.

On a :

$$NE^2 = (y_1 - \omega)^2 + \ldots + (y_N - \omega)^2$$

Parmi ces quantités, certaines correspondent à y'_1, y'_2, ..., d'autres non. Donc cette somme est supérieure à la somme $(y'_1 - \omega)^2 + \ldots\ldots + (y'_n - \omega)^2$. Chacune d'elles est supérieure à t^2E^2. Donc $NE^2 \geqq nt^2E^2$.

En divisant par E^2, on obtient : $N \geqq nt^2$, d'où : $\frac{n}{N} \leqq \frac{1}{t^2}$.

Mais $\frac{n}{N}$ est le quotient du nombre de celles des épreuves où la condition est vérifiée, par le nombre

total des épreuves. Donc, c'est la fréquence cherchée F_t, et on peut écrire

$$F_t \leq \frac{1}{t^2}$$

Par exemple la fréquence des écarts individuels qui sont $\geqq$ 3 fois l'écart quadratique moyen est $\leqq \frac{1}{9}$.

Ceci, pour un nombre arbitraire w, permet de mesurer pour ainsi dire la dispersion des y à partir de w. On voit en outre que non seulement cette dispersion dépend de t, mais aussi de E^2, et par conséquent de w. La dispersion la plus faible s'observera autour de la valeur w pour laquelle E^2 est le plus petit. C'est ce cas seulement qui semble avoir attiré l'attention de Bienaymé et de Tchebichef : celui pour lequel E^2 est minimum.

Cherchons maintenant la quantité w qui rend minimum l'écart quadratique moyen ainsi défini.

Nous commencerons par énoncer une règle utile, en nous appuyant sur la formule bien connue :

$$(a \pm b)^2 = a^2 \pm 2ab + b^2$$

Cette règle est la suivante : le carré de l'écart quadratique moyen E des y avec un nombre w est égal à la somme du carré de l'écart quadratique moyen μ des y avec leur moyenne arithmétique m et du carré de l'écart de w avec m.

En effet :

$$(y_1 - w) = (y_1 - m) + (m - w), \text{ d'où}$$
$$(y_1 - w)^2 = (y_1 - m)^2 + (m - w)^2 + 2(y_1 - m)(m - w)$$

et

$$(y_1 - w)^2 + \ldots + (y_N - w)^2 = (y_1 - m)^2 + \ldots + (m - w)^2 + \ldots + 2(m - w)(y_1 - m) + \ldots ;$$

d'où

$$E^2 = \frac{(y_1 - w)^2 + \ldots}{N} = \frac{(y_1 - m)^2 + \ldots}{N} + (m - w)^2 + 2(m - w)\left[\frac{(y_1 - m) + \ldots}{N}\right]$$

Or la moyenne arithmétique m de y_1, y_2,... y_N est

$$m = \frac{y_1 + y_2 + \ldots + y_N}{N}, \text{ d'où :}$$

$$y_1 - m + y_2 - m + \ldots = Nm - Nm = 0 ;$$

et

$$\mu^2 = \frac{(y_1 - m)^2 + \ldots + (y_N - m)^2}{N} ;$$

donc

$$E^2 = \mu^2 + (m - w)^2 \tag{15}$$

Comme conséquence : si j'appelle μ^2 la valeur de E^2 quand on remplace w par m, c'est-à-dire l'écart quadratique moyen des y avec leur valeur moyenne m, on voit que

$E^2 = \mu^2 +$ une quantité positive. Donc $E^2 \geqq \mu^2$, et $E \geqq \mu$.

Le minimum de l'écart quadratique moyen des y avec un nombre variable w est donc obtenu quand w est égal à la moyenne arithmétique m des y.

Application. — Reprenons l'exemple du nombre mensuel des jours de pluie en 1917, 1918 et 1919 :

Calculons l'erreur quadratique moyenne à partir de 14, pour 1917 et 1918, et de 15 pour 1919. Comme ce sont les valeurs médianes, ces erreurs quadratiques, sans être minima, seront voisines des minima.

En faisant la somme des carrés des écarts mensuels et en la divisant par 12 pour chaque année, on obtient :

En 1917 :

$$E_1 = \sqrt{\frac{191}{12}} = \sqrt{15{,}9}\,;$$

En 1918 :

$$E_2 = \sqrt{\frac{271}{12}}\,;$$

En 1919 :

$$E_3 = \sqrt{\frac{175}{12}}\,.$$

Cherchons maintenant la valeur moyenne des y = somme des nombres mensuels divisée par 12 :

En 1917 :

$$M = \frac{177}{12} = 14{,}75\,;$$

En 1918 :

$$M = \frac{175}{12} = 14{,}58\,;$$

En 1919 :

$$M = \frac{183}{12} = 15{,}25.$$

Calculons les écarts quadratiques moyens correspondants μ_1, μ_2, μ_3. Le plus commode est de chercher ces écarts à partir d'une valeur voisine de la moyenne arithmétique, mais présentant moins de décimales,

14 pour 1917 et 1918, et 15 pour 1919, par exemple. On a alors en vertu de la formule (15) :

$$E_1^2 = \mu_1^2 + (14,75 - 14)^2 ; \quad E_2^2 = \mu_2^2 + (14,58 - 14)$$
$$E_3^2 = \mu_3^2 + (15,25 - 15)^2.$$

D'où

$$\mu_1^2 = \frac{191}{12} - (0,75)^2 = 15,9167 - 0,5625 = 15,35 ;$$
$$\mu_2^2 = \frac{271}{12} - (0,58)^2 = 22,47 ; \quad \mu_3^2 = \frac{475}{12} - (0,25)^2 = 39,52.$$

$\mu_1 = 3,9$; $\mu_2 = 4,7$; $\mu_3 = 6,3$, quantités peu différentes de E_1, E_2, E_3. On voit que ces écarts quadratiques moyens sont respectivement supérieurs aux écarts moyens qu'on a trouvés égaux respectivement à 2,91 ; 3,91 ; 5,25. Ce n'est qu'un cas particulier d'une remarque plus générale :

L'écart quadratique moyen de plusieurs valeurs y à partir d'un nombre quelconque ω est au moins égal à l'écart moyen de ces valeurs à partir du même nombre ω.

En effet appelons α, β, ... λ les écarts en valeurs absolues des diverses valeurs de y à partir de ω. L'écart moyen sera : $\varepsilon = \dfrac{\alpha + \ldots\ldots + \lambda}{N}$, et l'écart quadratique moyen : $E = \sqrt{\dfrac{\alpha^2 + \ldots\ldots + \lambda^2}{N}}$, N étant le nombre des épreuves.

Je vais montrer que $\varepsilon^2 \leqq E^2$, ou que $N^2\varepsilon^2 \leqq N^2E^2$.

On a :

$$N^2\varepsilon^2 = (\alpha + \ldots + \lambda)^2$$

et

$$N^2E^2 = N(\alpha^2 + \ldots + \lambda^2)$$

$$N^2E^2 - N^2\varepsilon^2 = (\alpha^2 + \ldots + \lambda^2)N - (\alpha + \ldots + \lambda)^2$$

Or on a :

$$(\alpha + \beta + \ldots + \lambda)^2 = \alpha^2 + \beta^2 + \ldots \lambda^2 + 2\alpha\beta + \ldots$$

Donc

$$N^2E^2 - N^2\varepsilon^2 = (\alpha^2 + \ldots + \lambda^2)N - (\alpha^2 + \ldots + \lambda^2 + 2\alpha\beta + \ldots) = (N-1)\alpha^2 + (N-1)\beta^2 + \ldots - 2\alpha\beta - \ldots$$

Or ceci est égal à

$$(\alpha - \beta)^2 + (\alpha - \gamma)^2 + \ldots + (\beta - \gamma)^2 + \ldots \geqq 0.$$

Donc $E^2 \geqq \varepsilon^2$.

Puisque l'écart quadratique moyen à partir de ω est au moins égal à l'écart moyen à partir de ω, et que celui-ci est au moins égal à l'écart moyen tout court (c'est-à-dire à partir de la valeur médiane), on voit que l'écart quadratique moyen à partir de ω est au moins égal à l'écart moyen tout court. Et par suite, enfin, l'écart quadratique moyen tout court c'est-à-dire à partir de la moyenne arithmétique est au moins égal à l'écart moyen tout court.

Utilité des valeurs typiques. — Pour montrer l'utilité des valeurs typiques nous prendrons une statistique dressée par Pearson[1], pour mettre en évidence la

1. *Philosophical Transactions*, A, vol. CXCII, 1899.

covariation entre le nombre des enfants de la mère et de la fille. On a relevé les nombres des enfants de mille pairesses anglaises et les nombres des enfants d'une de leurs filles.

Le problème posé, c'est d'abord de savoir s'il y a une relation entre deux nombres, ensuite de déterminer cette relation. Si ces deux nombres correspondaient à deux valeurs physiques liées par une loi simple, à toute valeur de l'une on pourrait faire correspondre une valeur de l'autre : par exemple, la pression sur le fond d'un vase, à la hauteur de la colonne d'eau. Si on dresse le graphique correspondant, en marquant des points dont les distances à deux droites fixes ont respectivement pour mesure la hauteur et la pression, on voit immédiatement sur ce graphique comment la pression varie en fonction de la hauteur.

Mais, dans le problème actuel, on ne peut plus opérer ainsi. Appelons M le nombre des enfants de la mère, et F le nombre des enfants de la fille. A chaque valeur de M correspondent plusieurs valeurs de F. Et, quel que soit M, on trouve des mêmes F : aussi bien parmi les filles uniques que parmi les filles appartenant à une nombreuse famille on trouvera des mères de 1, 2, 3, 4,..... enfants. Il faut transformer le problème pour être en mesure d'étudier la relation, si l'on admet qu'il en existe une. Il faut faire correspondre à chaque valeur de M un seul nombre F alors qu'il y en a plusieurs ; on est donc amené à faire correspondre à chaque valeur de M un nombre qui sera choisi de façon à représenter

l'ensemble des valeurs correspondantes de F ; c'est-à-dire une valeur typique F_M de cet ensemble de valeurs F. On sera ainsi ramené à un problème du genre précédent : variation de F_M en fonction de M.

Appliquons ce que nous venons de dire à l'exemple de Pearson. Celui-ci est précisé par le tableau suivant, qui indique le nombre de cas où la mère avait M enfants et la fille F enfants.

Première ligne = M

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	11-16	Totaux	M_F
0	5	9	11	18	21	15	9	9	6	3	2	3	—	—	—	—	5	110	5
1	12	5	14	15	10	13	9	8	5	3	2	2	—	—	—	—	4	98	5
2	9	9	10	15	18	15	13	3	2	4	2	—	—	—	—	1	3	97	5
3	5	10	16	11	9	14	18	10	4	8	2	3	—	—	—	—	5	105	6
4	5	5	19	17	21	15	18	10	14	2	1	5	1	—	—	—	7	133	5
5	7	6	7	17	23	9	15	13	14	8	3	2	2	—	—	—	7	123	6
6	4	5	8	11	15	12	9	14	7	5	3	3	1	—	—	—	7	103	6
7	5	4	3	8	4	13	8	8	5	10	2	1	1	—	—	—	4	73	6
8	1	2	4	12	9	9	7	5	12	3	4	1	2	1	—	—	8	73	6
9	—	—	4	3	3	4	4	5	3	2	2	1	—	—	—	—	3	34	7
10	—	—	1	2	1	3	—	6	3	2	—	1	—	1	—	—	2	24	8
11	—	—	2	1	1	1	1	—	1	2	—	—	—	—	—	—	0	8	5 1/2
12	—	2	1	2	3	—		1	—	—	1	—	1	—	1	—	3	13	5
13	—	—	—	—	2	1	—	—	—	—	1	—	2	—	—	—	3	6	8 1/2
Totaux	53	57	100	132	140	124	113	92	76	52	25	22	10	2	1	1	61	1.000	6
F_M	3	3	3	4	4	4	4	5	5	5	6	4	7 1/2	9	12	2	5	4	

Le tableau a été complété par les indications des valeurs médianes M_F (correspondant aux lignes), et F_M (aux colonnes).

Première colonne = F

On a choisi par exemple comme valeur typique la médiane, plus facile à calculer.

Soit 0, 1, 2..... les valeurs F. Je cherche les répétitions de ces valeurs, dans la colonne 1, par exemple pour calculer F_1. O est répété r_0 fois $= 5$; 1, r_1 fois $= 12$;..... On fait la somme des nombres r_0, r_1,..... jusqu'à ce qu'on atteigne et dépasse la moitié du nombre total des cas. Par exemple $r_0 = 5$; $r_0 + r_1 = 17$; $r_0 + r_1 + r_2 = 26 < \frac{53}{2}$; $r_0 + r_1 + r_2 + r_3 = 31 > \frac{53}{2}$; donc 3 est la valeur médiane F_1 de F quand $M = 1$. Les nombres F_M ainsi obtenus vont en croissant, mais jusqu'à un certain point seulement, où les observations deviennent rares. La régularité de la croissance jusqu'à ce point est manifeste. On voit bien que, d'une façon générale, les enfants de familles nombreuses ont eux-mêmes de nombreux enfants.

Comparaison des valeurs typiques. — Comparons les avantages et les inconvénients respectifs de la moyenne et de la médiane.

Elles satisfont toutes deux à certaines conditions auxquelles doit répondre toute valeur typique pour jouer le rôle qu'on lui assigne.

1° Si en changeant d'unité de mesure, les nombres y_1, y_2, ... y_N sont multipliés par un même facteur la valeur typique du nouvel ensemble de nombres doit être égale à l'ancienne multipliée par ce même facteur. De la sorte, il y aura une grandeur typique d'un ensemble

de grandeurs, indépendamment de l'unité de mesure.

2° Si le groupe de valeurs est formé de valeurs toutes égales à un certain nombre, la valeur typique doit avoir la valeur de ce nombre.

3° Si on a un groupe de valeurs $y_1 \leqq y_2 \ldots\ldots \leqq y_N$, une valeur typique de cet ensemble doit être comprise entre la plus grande et la plus petite.

4° Supposons que je cherche les valeurs typiques de plusieurs groupes, dont chacun représente les points obtenus dans une série de coups de dés, par exemple un groupe g de 300 coups de dés, un groupe g' de 100, etc. Supposons que le groupe g soit constitué par la répétition de trois groupes tels que g'. Les fréquences de chaque point seront les mêmes dans les deux groupes. Il est évident que les valeurs typiques devront être les mêmes. Ainsi, les fréquences étant données, les valeurs typiques devront être connues. En d'autres termes si on a un groupe de grandeurs qui prennent des valeurs distinctes x_0, $x_1, \ldots\ldots x_n$, et si f_0, $f_1, \ldots\ldots f_n$ sont les fréquences de ces valeurs dans le groupe, l'ensemble de ces fréquences doit définir la valeur typique. — En généralisant cette condition, on obtient la proposition suivante.

5° Soient deux groupes h et h', dans lesquels les fréquences d'une même valeur sont différentes. Si les fréquences correspondantes diffèrent peu, les valeurs typiques doivent être peu différentes.

Ces conditions sont-elles vérifiées, et au même degré, par la moyenne arithmétique et la médiane ?

Il en est bien ainsi pour la condition 1°.

Si $y_0 = y_1 = \ldots\ldots = y_N$, la moyenne et la médiane sont la valeur commune [condition (2°)].

Si $y_0 \leqq y_1 \ldots\ldots \leqq y_N$, la condition (3°) est évidemment vérifiée par la médiane qui est une des valeurs de y. Pour la moyenne v, on a $Nv = y_1 + \ldots\ldots + y_N$. Donc $\frac{Nv}{N} \leqq y_N$ et $\geqq y_1$. Donc $y_1 \leqq v \leqq y_N$.

La condition (4°) est vérifiée pour la moyenne. En effet la moyenne arithmétique sera de la forme : $v = \frac{r_1 x_1 + r_2 x_2 + \ldots\ldots + r_n x_n}{N}$, $r_1, r_2 \ldots\ldots$ désignant le nombre de fois que $x_1, x_2, \ldots\ldots$ se retrouvent dans le groupe d'épreuves $y_1, y_2, \ldots\ldots$ Mais nous savons que $\frac{r_1}{N}$, c'est la fréquence de x_1 dans le groupe. Donc $v = f_1 x_1 + f_2 x_2 + \ldots\ldots + f_N x_N$. Elle est la même pour deux groupes où les fréquences sont les mêmes. — Cela est moins évident pour la médiane. Comment la calcule-t-on, connaissant les fréquences? Si N est le nombre d'épreuves, il faudra calculer les sommes r_1 ; $r_1 + r_2$;..... ou $f_1 N$; $f_1 N + f_2 N$;..... jusqu'à ce qu'on arrive à une somme qui atteigne et dépasse $\frac{N}{2}$. Cela revient à former les sommes des fréquences f_1 ; $f_1 + f_2$; ..... jusqu'à ce qu'on atteigne et dépasse $\frac{1}{2}$. Il n'intervient dans ce calcul que les x et les f. Donc, si les f sont égaux dans les deux groupes, la médiane sera la même.

Mais il y a d'autres propriétés qui n'appartiennent

pas au même degré à la moyenne arithmétique et à la médiane. Une condition intéressante est la condition 5° : il y a avantage à définir la valeur typique d'un groupe de grandeurs de telle façon qu'elle change très peu quand les fréquences varient très peu. A ce point de vue, la moyenne est supérieure à la médiane. Supposons qu'on cherche l'âge typique d'une certaine population en 1912 et en 1913. Soit $f_1, f_2, \ldots\ldots f'_1, f'_2, \ldots\ldots$ les fréquences de ces âges. La moyenne sera : $0\,f_0 + 1\,f_1 + 2\,f_2 + \ldots\ldots$ en 1912, et $0\,f'_0 + 1\,f'_1 + f'_2 + \ldots\ldots$ en 1913. Si les fréquences sont voisines, f_0 voisin de f'_0, f_1 de f'_1 ....., les sommes différeront très peu. C'est ce qu'on exprime en disant que la moyenne est une fonction continue des fréquences. Dire, d'autre part, que 55 est l'âge médian, c'est dire que, si on fait les sommes f_0 ; $f_0 + f_1$; ..... on arrive à une dernière somme $f_0 + f_1 + \ldots\ldots + f_{54}$ inférieure à $\frac{1}{2}$. Or il peut se produire que cette somme soit à peine inférieure à $\frac{1}{2}$, et qu'en variant très peu elle lui devienne supérieure : on passerait donc brusquement de 55 à 54. C'est pourquoi on emploie fréquemment, en statistique, les moyennes arithmétiques, quand on veut suivre d'une façon continue les valeurs que peut prendre une grandeur donnée. — Remarquons toutefois que ce défaut de la médiane est plus apparent que réel ; il peut offrir cet avantage de rappeler que le choix de la valeur typique est un peu arbitraire.

Une condition (qui ne s'impose pas immédiatement

pour le choix d'une valeur typique, est que le calcul de la valeur typique ne soit pas trop compliqué. D'une façon générale, la valeur moyenne peut se calculer plus facilement et a un sens plus intuitif, quand il s'agit d'une grandeur homogène qui s'additionne, se divise ou se répartit, comme de l'argent, du vin. Par exemple, la consommation moyenne par tête d'habitant : il serait pratiquement impossible de connaître d'abord et de calculer ensuite la consommation de chaque habitant, ce qui serait nécessaire pour calculer la médiane, alors que, pour la moyenne, il suffit de connaître la consommation totale. Il n'en est pas de même de grandeurs telles que l'âge, la taille, etc. : ici, la représentation d'une échelle, d'une série continue de variations individuelles est plus naturelle, l'âge médian, la taille médiane sont des nombres plus typiques.

En outre dans tous ces cas le calcul de la médiane est plus rapide que celui de la moyenne. Par exemple, d'après le tableau de la page 123, je calcule les médianes F′ et les moyennes F″ correspondant aux colonnes : nombre d'enfants d'une fille dont la mère avait un nombre d'enfants M donné.

M	1	2	3	4	5	6	7	8	9	10	11-16
F′	3	3	3	4	4	4	4	5	5	5	5
F″	3,15	3,33	3,86	3,86	4,11	4,09	4,16	4,69	4,85	4,98	5,72

Pour calculer les médianes, il m'a suffi de faire 11 additions, et 11 divisions par 2.

Pour calculer les moyennes arithmétiques, j'ai dû

faire 13 multiplications (à peu près) pour chacune des 11 colonnes, soit une centaine au moins, puis 11 additions, puis 11 divisions (le plus souvent par des nombres de deux chiffres).

Lorsqu'on cherche la valeur typique d'un groupe, l'origine concrète de ces groupes peut être différente, et il s'ensuit que la moyenne et la médiane pourront ne pas répondre aux mêmes conditions. Si je cherche un nombre qui représente des grandeurs mesurables, telles que des sommes d'argent, la moyenne convient. Si, d'autre part, je cherche qu'elle est la couleur typique des cheveux dans une population donnée, je pourrai prendre des échantillons d'une couleur déterminée, que je numéroterai du plus pâle au plus foncé, et je désignerai par r_1 l'ensemble de ceux qui ont des cheveux dont la couleur se rapproche de l'échantillon 1, r_2, de 2, etc. Je peux dégager la valeur typique de ces nombres, de façon à établir des comparaisons entre diverses populations. Mais les nombres 1, 2, etc., ne servent qu'à *repérer*, et non à mesurer directement : ils n'ont pas une valeur absolue. Supposons que je calcule la moyenne : $\frac{1r_1 + 2r_2 + \text{etc.}}{r_1 + r_2 + \text{etc.}} = v$, et que je trouve un nombre compris entre 6 et 7 : cela signifie que la couleur typique est comprise entre les deux échantillons numérotés 6 et 7. Mais mon numérotage a été arbitraire. J'aurais pu dire : 0, 1, 2, ... au lieu de 1, 2, 3,..... et trouver pour v un nombre différent correspondant à une teinte typique différente. — Ou encore, si je compare des draps d'un gris plus ou

moins intense, obtenu avec une teinture noire mélangée à de l'eau, je peux soit numéroter les teintes, soit calculer le poids de teinture utilisé pour chacun d'eux : suivant le procédé que je choisirai j'obtiendrai, pour v, autant de valeurs différentes qui correspondront, et c'est là l'inconvénient, à des teintures moyennes différentes. Au contraire la valeur médiane, quelle que soit la convention adoptée, et pourvu qu'on respecte l'ordre d'intensité, correspondra toujours à un même échantillon médian. — Par exemple encore, si je cherche, parmi un certain nombre de carrés, celui qui est typique au point de vue grandeur, j'obtiendrai des résultats différents, si je cherche celui qui a un côté d'une longueur moyenne, ou une surface d'une étendue moyenne. Soient 5 carrés de côté = 84 ; 86 ; 100 ; 101 ; 128 (en dcm) : la longueur moyenne du côté est $v = \frac{499}{5} = 99{,}8$, qui est plus petit que le côté 100 du 3ᵉ carré. Leurs surfaces sont respectivement = 7.056 ; 7.396 ; 10.000 ; 10.201 ; 16.384 (en dcmq) ; la moyenne est : $v = \frac{51.037}{5} = 10.207{,}4$, ce qui est plus grand que la surface du 4ᵉ carré. Il y a donc au moins deux carrés moyens, l'un entre le 2ᵉ et le 3ᵉ carrés donnés, l'autre entre le 4ᵉ et le 5ᵉ, selon qu'on chiffre la grandeur d'un carré par la longueur de son côté ou par son aire. Au contraire il est évident qu'il n'y a qu'un carré médian, qui est le 3ᵉ pour le groupe donné.

Combinaison des valeurs typiques. — Nous pou-

vons avoir à chercher la moyenne d'un groupe formé de plusieurs groupes, alors que nous connaissons, non pas la répartition des fréquences, mais la moyenne arithmétique, pour chacun d'eux. Connaissant ces moyennes des groupes composants, nous trouvons sans peine la moyenne de l'ensemble.

Soient ces moyennes :

$$v = \frac{y_1 + y_2 + \dots + y_N}{N}\ ;\quad v' = \frac{y'_1 + y'_2 + \dots + y'_{N'}}{N'}\ ;$$

$$V = \frac{y_1 + y_2 + \dots + y_N + y'_1 + y'_2 + \dots + y'_N}{N + N'} = \frac{Nv + N'v'}{N + N'}$$

Cela signifie que la moyenne dans le groupe total est celle qu'on obtiendrait, si, dans le premier groupe, toutes les valeurs étaient égales à leur moyenne, et de même dans le deuxième ; cela rend le calcul très commode. — La formule peut se généraliser, au cas où l'on a un nombre quelconque de groupes :

$$V = \frac{N_1 v_1 + N_2 v_2 + \dots}{N_1 + N_2 + \dots}$$

Premier exemple. — Si la moyenne de la taille prend les valeurs 1m,62 ; 1m,63 ; 1m,64 ; 1m,65 dans quatre compagnies de 250, 200, 300 et 400 hommes, la taille moyenne dans le bataillon sera le quotient :

$$\frac{1{,}62 \times 250 + 1{,}63 \times 200 + 1{,}64 \times 300 + 1{,}65 \times 400}{250 + 200 + 300 + 400} = 1^{m}{,}637.$$

Deuxième exemple. — Soit à calculer la durée moyenne de vie en France. Les tables donnent le nombre de survivants par âge pour un groupe de 100.000 personnes. Il y a r_0 nouveau-nés, r_1 survivants à un an, etc. Si ces

personnes avaient, pour chaque nombre d'années, exactement cet âge, il suffirait de considérer les produits : $0 \times r_0 + 1 \times r_1 + \ldots\ldots + 30 \times r_{30} + \ldots\ldots$, et de diviser cette somme par la population de la France. Mais en réalité les personnes de chaque nombre d'années ont vécu ce nombre d'années $+$ une fraction. Admettons alors que l'âge moyen des personnes de 30 ans par exemple est 30 ans $\frac{1}{2}$. Et appliquons notre théorème. La durée moyenne de vie de la population sera alors égale à la vie moyenne d'un ensemble de groupes dans lesquels l'âge exact serait remplacé par l'âge moyen du groupe. Quelle est l'erreur commise? — En appelant $0 + \varepsilon_0$, $1 + \varepsilon_1$ ..... les âges moyens exacts des personnes de 0 an, 1 an, ..., et N la population totale, l'âge moyen de cette population est :

$$M = \frac{(0 + \varepsilon_0)\, r_0 + (1 + \varepsilon_1)\, r_1 + \ldots + (30 + \varepsilon_{30})\, r_{30} + \ldots}{N}$$

auquel on a substitué :

$$M' = \frac{\frac{1}{2}\, r_0 + \left(1 + \frac{1}{2}\right) r_1 + \ldots + \left(30 + \frac{1}{2}\right) r_{30} + \ldots}{N}$$

La différence est :

$$M' - M = \frac{\left(\frac{1}{2} - \varepsilon_0\right) r_0 + \left(\frac{1}{2} - \varepsilon_1\right) r_1 + \ldots}{N}$$

Or l'âge moyen des personnes de x ans est entre x et $x + 1$; donc ε_x est au plus une année. Donc $\left(\frac{1}{2} - \varepsilon_x\right)$ est compris entre $-\frac{1}{2}$ et $+\frac{1}{2}$ années. Donc :

$$|\,M - M'\,| \leqq \frac{\frac{1}{2}r_0 + \frac{1}{2}r_1 + \dots}{N} = \frac{1}{2} \cdot \frac{r_0 + r_1 + \dots}{N} = \frac{1}{2} \cdot \frac{N}{N} = \frac{1}{2}$$

L'erreur commise sera au plus d'une demi-année, et en fait elle sera beaucoup plus petite.

Il en sera de même si on veut calculer les moyennes des tailles d'une population. On part de mesures exactes à 1 centimètre près, et on arrive à un résultat dans lequel l'erreur commise est certainement très inférieure à 5 millimètres, c'est-à-dire à une précision plus grande que celle des observations.

— Dans le cas des médianes, c'est moins simple. Nous connaissons les médianes m_1, m_2, ... m_p des groupes de valeurs g_1, g_2, ... g_p, et nous voulons calculer la médiane M de l'ensemble des groupes. Dans le cas général, on ne peut procéder comme pour la moyenne dans le cas précédent : on n'arriverait pas à un résultat précis. Dans certains cas cependant, le calcul peut se simplifier de la même manière. On pourra, sans altérer M, remplacer tous les nombres de l'un des groupes par la valeur médiane correspondante, lorsque tous ces nombres seront inférieurs à M. Supposons en effet tous les nombres du groupe total rangés par ordre de valeurs croissantes, $z_1 < z_2 < \dots\dots$; M sera un de ces nombres $= z_n$.

Si tous les nombres du groupe g_1 sont inférieurs à M, le nombre m_1, valeur médiane de g_1, sera aussi inférieur à M. Si je remplace les nombres de g_1 par m_1, j'aurai changé l'ordre des z ; mais ceux qui étaient plus

petits que M le restent, et M reste le terme médian de l'ensemble.

Soit le cas particulier où tous les nombres de $g_1 \leqq$ tous les nombres de $g_2 \leqq$ tous les nombres de g_3, etc. Alors : $m_1 \leqq m_2 \leqq m_3$, ... Essayons de calculer M, connaissant seulement les m, et le nombre des épreuves de chaque groupe. Appelons M′ la médiane du groupe total obtenu en remplaçant tous les nombres de chaque groupe par la médiane de ce groupe. Soit : $m_1 \leqq m_2 \leqq m_3 \ldots \leqq m_{k-1} \leqq m_k$, etc. D'après ce qui précède, M est aussi la médiane de l'ensemble obtenu en remplaçant dans chaque groupe tous les nombres par leur médiane, sauf (peut-être) dans le groupe g_k qui comprend M. Par conséquent on a : $m_{k+1} \geqq M \geqq m_{k-1}$. Par suite, la différence M — M′ sera inférieure à $(m_{k+1} - m_{k-1})$ en valeur absolue. Si les médianes des groupes successifs varient peu, M sera peu différent de M′. Si elles diffèrent de moins de ε, $| M - M' | \leqq \varepsilon$.

Objets ou individus typiques. — Etant donné un groupe d'objets ou d'individus, on peut être tenté de les représenter par un objet ou un individu typique de la manière suivante : à tout caractère mesurable ou repérable de ces individus, on ferait correspondre la valeur typique des mesures ou indices de ce caractère. Par exemple on pourra appeler triangle rectangle moyen, dans un groupe de triangles rectangles, celui qui aurait pour côtés de l'angle droit la moyenne des bases et la moyenne des hauteurs. Le statisticien belge

Quetelet définissait un homme moyen qui aurait pour taille, pour poids, etc., la moyenne des tailles, des poids, etc., de l'ensemble des hommes.

Cournot, puis Bertrand ont fait à cette conception une objection simple et apparemment décisive : elle n'a de sens que si les diverses valeurs moyennes envisagées sont compatibles. Reprenons l'exemple des triangles rectangles. La valeur moyenne des angles droits serait un angle droit ; l'hypoténuse moyenne serait la moyenne des hypoténuses. Or si on prend deux triangles rectangles ayant l'un pour base 4 mètres ; pour hauteur 3 mètres ; pour hypoténuse 5 mètres ; l'autre pour base 3 mètres ; pour hauteur 4 mètres ; pour hypoténuse 5 mètres, le triangle rectangle moyen aura pour base $3^{m},5$; pour hauteur $3^{m},5$; pour hypoténuse, 5 mètres. On devrait donc avoir $(3,5)^2 + (3,5)^2 = 5^2$ ce qui est faux.

A plus forte raison, dans le cas complexe de l'homme moyen, arriverait-on à une foule d'incompatibilités.

L'emploi des valeurs médianes conduirait aux mêmes contradictions.

Il serait, cependant, plus satisfaisant en général à d'autres égards, dans les cas où on a un très grand nombre d'observations d'une grandeur continue.

En effet la valeur médiane n'aurait plus alors l'inconvénient, qu'elle présente d'ordinaire, de varier de façon discontinue. Et on aurait par elle une idée plus juste de la valeur typique. Par exemple, s'il s'agit de la fortune ou du revenu individuel dans une grande nation, on

calcule d'ordinaire la moyenne : mais la valeur obtenue n'est pas typique, puisqu'elle reflète dans une plus large mesure les plus grosses fortunes qui sont très rares. La valeur médiane représente mieux le type des fortunes trouvées dans l'ensemble.

Comparaison des écarts. — Pearson appelle l'écart quadratique moyen : « standard deviation », comme s'il était le seul qu'il faille communément employer. Si, dans les recherches théoriques, il faut préférer l'écart quadratique moyen, c'est qu'il possède des propriétés mathématiques simples, et qu'il est d'un emploi plus commode.

Mais les mêmes raisons de commodité conduisent à choisir au contraire l'écart moyen, dans les calculs numériques. C'est pourquoi, lorsque le mathématicien a simplifié l'exposé théorique en introduisant la notion d'écart quadratique moyen, il doit se plier aux exigences du calculateur, et indiquer comment, dans les applications, utiliser l'écart moyen, qui est d'un emploi plus aisé, et ne nous éloigne pas toutefois sensiblement des résultats de la théorie[1].

Calcul numérique des valeurs typiques et des écarts. — Proposons-nous de montrer sur un exemple comment il est avantageux de disposer les calculs pour

1. Remarquons à cette occasion que la notion d'écart moyen, liée à celle de valeur médiane, ne partage pas le plus grand inconvénient de celle-ci, savoir d'être, dans certains cas, indéterminée entre deux nombres.

déterminer numériquement les valeurs typiques et les écarts d'un groupe de nombres.

Il est d'abord indispensable pour le calcul de la valeur médiane, et il est avantageux pour le calcul de la valeur moyenne et des écarts, de ranger les nombres obtenus par ordre de grandeur, en indiquant pour chacun d'eux sa fréquence.

Par exemple, Czuber a observé les fréquences d'un certain événement fortuit dont nous reparlerons (p. 210), et il a dressé un tableau donnant les résultats de ses observations dans l'ordre où il les a obtenus.

Rang de l'observation	Valeur corresp. de la fréq. *f*.	Rang de l'observation	Valeur corresp. de la fréq. *f*.
1	0,17	20	0,16
2	0,20	21	0,15
3	0,10	22	0,15
4	0,17	23	0,18
5	0,12	24	0,16
6	0,15	25	0,14
7	0,19	26	0,18
8	0,22	27	0,17
9	0,17	28	0,14
10	0,19	29	0,13
11	0,14	30	0,15
12	0,22	31	0,22
13	0,18	32	0,19
14	0,17	33	0,16
15	0,13	34	0,26
16	0,12	35	0,14
17	0,18	36	0,15
18	0,15	37	0,14
19	0,17		

Et il opère directement sur cette table. Mais si on cherche les valeurs moyennes et les écarts moyens, le rang d'observation n'a pas d'intérêt. Il est préférable de ranger les fréquences par ordre de grandeur. Pour cela on pourrait être tenté de lire chaque nombre successivement, et de tâcher de les placer par ordre. Il vaudra mieux, d'ordinaire, écrire d'avance la liste des valeurs successives que peuvent prendre les nombres du groupe, et se contenter d'inscrire à côté leur répétition, c'est-à-dire le nombre r de fois qu'elles se présentent. Une inspection rapide montre par exemple ici que f reste compris entre 0,10 et 0,26 ; on écrit alors 0,10 ; 0,11 ; etc..., 0,26. On aura ainsi le tableau suivant :

f.	r.	f.	r.
0,10	1	0,20	1
0,11	0	0,21	0
0,12	2	0,22	3
0,13	2	0,23	0
0,14	5	0,24	0
0,15	6	0,25	0
0,16	3	0,26	1
0,17	6		
0,18	4	TOTAL	37
0,19	3		

On voit tout de suite que la plupart des observations donnent des valeurs de f voisines de 0,14 — 0,17.

Pour trouver exactement la médiane, on ajoute successivement les valeurs de r à partir de la gauche (ou de la droite) jusqu'à ce qu'on atteigne la moitié, $\frac{37}{2}$ $= 18\ 1/2$; du nombre total des épreuves. On voit que $1 + 0 + 2 + 2 + 5 + 6 = 16$, et que si on ajoutait le nombre de fois, 3, que se présente la valeur $f = 0,16$, on atteindrait et dépasserait $18\ 1/2$. Donc $0,16$ est la valeur médiane.

Calculons maintenant l'écart moyen. Pour cela nous allons écrire sous les lignes du tableau précédent les valeurs correspondantes x des écarts avec $f = 0,16$. Ces écarts sont positifs, négatifs ou nuls. Si l'on appelle r_x la répétition de x, c'est-à-dire le nombre de fois qu'il se présente, l'écart moyen, étant la moyenne arithmétique des valeurs de $|x|$, sera

$$\sigma = \frac{\Sigma\ |x|\ r_x}{37}$$

en indiquant par le symbole Σ qu'on fait la somme des termes obtenus en faisant varier x entre ses limites. On pourrait ne faire aucune distinction entre les x positifs et négatifs. Il sera cependant plus commode de faire à part les deux sommes correspondantes en vue du calcul de la valeur moyenne v des f.

En effet : on a $f = 0,16 + (f - 0,16)$; par suite, la valeur moyenne d'une somme étant égale à la somme des moyennes des termes, v est égal à la moyenne du nombre fixe $0,16$, moyenne qui est $0,16$, augmentée

de la moyenne v' du nombre f — 0,16 que nous avons appelé x. Or

$$v' = \frac{\Sigma\, x\, r_x}{37}$$

où les x sont pris avec leur signe.

Si donc on calcule séparément la somme A des produits xr_x où x est positif, et la somme B des produits xr_x où x est négatif, on voit qu'on aura :

$$(1)\ \sigma = \frac{A + B}{37}\ ;\ (2)\ v = 0{,}16 + \frac{A - B}{37}$$

ce qui justifie le calcul séparé de A et de B.

On aura alors le tableau suivant, qu'on devra supposer placé immédiatement à côté du tableau précédent et en correspondance avec lui :

x	$x r_x$	x	$x r_x$
— 0,06	— 0,06	+ 0,05	0
— 0,05	0	+ 0,06	+ 0,18
— 0,04	— 0,08	+ 0,07	0
— 0,03	— 0,06	+ 0,08	0
— 0,02	— 0,10	+ 0,09	0
— 0,01	— 0,06	+ 0,10	+ 0,10
0	0		
+ 0,01	+ 0,06	Totaux A =	0,55
+ 0,02	+ 0,08		
+ 0,03	+ 0,09	B =	0,36
+ 0,04	+ 0,04		

D'où, à la fois :

$$\sigma = \frac{0{,}55 + 0{,}36}{37} = 0{,}0246\ ;\ v = 0{,}16 + \frac{0{,}55 - 0{,}36}{37} = 0{,}16514\ldots$$

On remarquera que l'on réduit ainsi extrêmement l'influence des erreurs éventuelles de calcul, puisque la formule directe $v = \frac{\Sigma fr}{37}$ ferait intervenir des nombres f plus grands que les nombres x (écarts) de la formule (2), et qu'en outre le terme de v correctif à 0,16 se présente sous la forme d'une quantité petite $\frac{A - B}{37}$, où A et B se compensent presque.

En somme cette méthode consiste à ramener le calcul à effectuer sur les f à un calcul sur les différences numériques plus simples x des quantités f avec une moyenne provisoire, ici 16.

Enfin l'emploi de cette sorte de « moyenne provisoire », suivant l'expression de Charlier, simplifie aussi le calcul de l'écart quadratique moyen μ_0. D'après la définition (voir p. 115 et p. 120).

$$\mu_0{}^2 = \frac{\Sigma\, r\, (f - v)^2}{37}\,.$$

On voit qu'il faudrait introduire les différences $f - v$, où figureraient (puisque $v = 0,1651.....$) un grand nombre de décimales, qui, en outre, n'ont pas encore été calculées. On peut éviter ces deux inconvénients grâce à la formule (15) déjà démontrée, page 117 :

$$\mu^2 = \mu_0{}^2 + (w - v)^2$$

où μ^2 est l'écart quadratique moyen des f à partir d'une valeur arbitraire w que nous allons prendre ici égale à 0,16 ; μ^2 est donc la moyenne arithmétique des carrés des écarts, $x = f - 0,16$.

D'où

$$\mu^2 = \frac{\Sigma\, x^2 r_x}{37}$$

et

$$\mu^2_0 = \mu^2 - (v - 0,16)^2.$$

Comme $(v - 0,16)$ est petit $(v - 0,16)^2$ le sera encore plus, de sorte qu'en général μ_0^2 sera très voisin de μ^2, et on aura pu se contenter de porter au carré les quantités simples x, au lieu d'avoir d'abord à calculer (avant de les élever au carré) les quantités compliquées $(f - v)$.

On devrait faire suivre les tableaux précédents (et immédiatement à côté) du tableau suivant :

x^2	$x^2 r_x$	x^2	$x^2 r_x$
0,0036	0,0036	0,0009	0,0027
0,0025	0	0,0016	0,0016
0,0016	0,0032	0,0025	0
0,0009	0,0018	0,0036	0,0108
0,0004	0,0020	0,0049	0
0,0001	0,0006	0,0064	0
0	0	0,0081	0
0,0001	0,0006	0,01	0,0100
0,0004	0,0016		
		TOTAL	0,0385

Donc :

$$\mu^2 = \frac{0,0385}{37} = 0,001040\ldots$$

$$\mu_0^2 = 0,001040 - (0,0051\ldots)^2 = 0,001040 - 0,000026\ldots$$

$$\mu_0 = 0,0318$$

Charlier a indiqué un procédé de vérification qui peut être commode et qui est basé sur les formules :

$$(x+1)^2 = x^2 + 2x + 1$$
$$(x+1)^2 r_x = x^2 r_x + 2x r_x + r_x$$
$$(16) \qquad \Sigma (x+1)^2 r_x = \Sigma x^2 r_x + 2 \Sigma x r_x + \Sigma r_x$$

En général les différences entre les valeurs successives des x sont égales ; on appelle cette valeur commune l'unité d'intervalle ; désignons par X le nombre obtenu en divisant x par cette unité d'intervalle, de sorte que les différences entre les valeurs successives des X seront toutes égales à l'unité.

Par exemple, dans le cas actuel, $x = -0{,}06$; $-0{,}05$; ..... On prendra $X = 100\, x$, d'où

$$X = -6 ; -5 ; \ldots$$

Alors les quantités $\Sigma X^2 r_x$, ici $10.000\, \Sigma x^2 r_x = 385$; $\Sigma X r_x$, ici $100\, \Sigma x r_x = 100\, (A - B) = 19$; Σr_x, ici 37, auront été calculées ; les quantités $(X+1)^2$ (ici 10.000 fois le carré qui suit x^2) aussi ; il sera facile de calculer $\Sigma (X+1)^2 r_x$, et de vérifier si l'égalité (16) est satisfaite.

On a alors la disposition suivante, où on a permuté dans ce qui précède les lignes et les colonnes, pour que les nombres à additionner se présentent en colonnes verticales :

$f.$	r	X	rX (—)	rX (+)	X²	rX²	r(X+1)²
0,10	1	—6	—6		36	36	25
0,11	0	—5	0		25	0	0
0,12	2	—4	—8		16	32	18
0,13	2	—3	—6		9	18	8
0,14	5	—2	—10		4	20	5
0,15	6	—1	—6		1	6	0
0,16	3	0	0		0	0	3
0,17	6	1		6	1	6	24
0,18	4	2		8	4	16	36
0,19	3	3		9	9	27	48
0,20	1	4		4	16	16	25
0,21	0	5		0	25	0	0
0,22	3	6		18	36	108	147
0;23	0	7		0	49	0	0
0,24	0	8		0	64	0	0
0,25	0	9		0	81	0	0
0,26	1	10		10	100	100	121
TOTAUX	37		—36 = —B′	+55 = +A′		385	460

$$\sigma = \frac{\Sigma \mid x \mid r_x}{37} = \frac{\Sigma \mid X \mid r_x}{3700} = \frac{A'+B'}{3700}$$

$$\nu - 0{,}16 = \frac{\Sigma\, x r_x}{37} = \frac{\Sigma\, X r_x}{3700} = \frac{A'-B'}{3700}$$

$$A' + B' = 55 + 36 = 91$$
$$A' - B' = 55 - 36 = 19$$

$$\sigma = \frac{91}{3700} = 0{,}0246$$

$$\nu = 0{,}16 + \frac{19}{3700} = 0{,}1651$$

$$\mu^2 = \frac{\Sigma\, x^2 r_x}{37} = \frac{\Sigma\, X^2 r_x}{370.000} = \frac{385}{370.000} = 0{,}00104$$

$$\mu_0{}^2 = \mu^2 - (\nu - 0{,}16)^2 = 0{,}001040 - (0{,}0051)^2 = 0{,}00101$$

$$\mu_0 = 0{,}0318$$

Vérification :

$$460 = \Sigma r_x(X+1)^2 = \Sigma rX^2 + 2\Sigma rX + \Sigma r = 385 + 38 + 37 = 460$$

Remarque. — Dans ce qui précède nous avons choisi comme moyenne provisoire la médiane 16. Pour le calcul de la moyenne arithmétique et de l'écart quadratique moyen ce choix n'est pas indispensable ; l'essentiel est de prendre comme moyenne provisoire un nombre numériquement simple ou dont les différences avec les valeurs données soient simples, et qui paraît d'avance devoir être voisin de la moyenne à calculer.

EXERCICES

I (F). — Partant du tableau donné page 102, des hauteurs de pluie tombée chaque mois, calculer pour chacune des années considérées : la valeur médiane et l'écart moyen, la moyenne arithmétique et l'écart quadratique moyen.

II (F). — Calculer, par les méthodes qui ont servi à déterminer la vie probable, le montant probable m d'une succession en 1907 et en 1912, d'après le tableau de la page 11.

Calculer de même le montant moyen v d'une succession à ces deux époques, sachant que le montant total de toutes les successions déclarées a été, en millions de francs : 5.462 en 1907 et 5.577 en 1912.

III (F.) — On peut considérer que l'inégalité de la répartition des fortunes est d'autant plus prononcée que le montant probable m est plus petit par rapport au montant unique qu'on constaterait si les fortunes étaient également distribuées c'est-à-dire au montant moyen v. On peut donc considérer le rapport $\frac{v}{m}$ comme mesurant ou mieux repé-

rant l'inégalité de répartition des fortunes (ou tout au moins des successions déclarées). On remarquera que ce rapport est indépendant de l'unité monétaire adoptée et qu'il ne varie pas non plus si on passe d'une population à une autre plus ou moins nombreuse que la première, mais où les fréquences des actifs sont les mêmes que dans la première.

Calculer les valeurs de l'indice d'inégalité des successions, $\frac{v}{m}$ (qu'on vient de définir) en 1907 et en 1912.

2e Section.

Valeurs typiques et écarts d'un nombre aléatoire dans une catégorie infinie d'épreuves. — Les considérations développées précédemment sur la moyenne arithmétique et la valeur médiane s'appliquent à tout groupe fini déterminé d'épreuves, et sont indépendantes des lois du hasard. Soit, maintenant, une catégorie infinie d'épreuves dont chacune fournit un nombre y qui ne peut prendre qu'un nombre fini de valeurs distinctes (par exemple, dans une catégorie infinie de coups de dés, on peut amener à chaque fois un des nombres déterminés 1, 2,..... 6). Comment définir alors la moyenne arithmétique et la médiane de y ?[1]

1. Pour marquer la différence entre les groupes finis et les catégories infinies d'épreuves, nous emploierons ici les expressions :

Je prends un groupe (fini) G d'épreuves dans la catégorie C (l'ensemble infini des jets d'un dé). Pour ce groupe et les valeurs correspondantes, je peux définir comme précédemment une moyenne v, une médiane m, un écart moyen σ, un écart quadratique moyen μ, que j'affecte d'un indice : $v^{(G)}$,..... Je suppose maintenant que, dans la catégorie C, les valeurs y puissent prendre un nombre fini de valeurs distinctes x_1,..... x_n. Ces valeurs ont, dans le groupe G, une certaine fréquence $f_n^{(G)}$ par exemple pour la valeur x_n. Je suppose, en appelant y une quelconque des valeurs du groupe, qu'il y a une probabilité déterminée pour que y prenne la valeur x_1, x_2,....., soit p_1, p_2,..... C'est admettre (voir p. 8, loi du hasard) que $f_1^{(G)}$,..... $f_n^{(G)}$ vont pouvoir être considérés comme les valeurs approchées de p_1,..... p_n, dans la catégorie considérée, lorsque le nombre N d'épreuves du groupe G sera suffisamment grand. Autrement dit, cette sorte de valeur limite est maintenant indépendante du groupe G proprement dit, c'est-à-dire sera approximativement la même pour tous les groupes G, à condition que G soit pris au hasard dans la catégorie considérée, et qu'il devienne de plus en plus nombreux. Donc les quantités $v^{(G)}$, $m^{(G)}$, $\sigma^{(G)}$, $[\mu^{(G)}]^2$ sont des valeurs approchées, quand n est assez grand, de valeurs v, m, σ, μ^2, indépendantes du choix du groupe G dans la catégorie C. — En effet la valeur moyenne d'un groupe de nombres est déterminée quand

valeur moyenne (au lieu de : moyenne arithmétique) et : valeur probable (au lieu de : médiane).

on connaît les valeurs distinctes prises par ce nombre ainsi que leurs fréquences (sans connaître le nombre des épreuves). La valeur moyenne, $v^{(G)} = f_1^{(G)} \times x_1 + f_2^{(G)} \times x_2 + \ldots\ldots + f_n^{(G)} \times x_n$. Si G est assez nombreux, $f_1^{(G)}, \ldots\ldots f_n^{(G)}$ seront voisins de $p_1, \ldots\ldots p_n$, et par conséquent $f_1^{(G)} \times x_1 + \ldots\ldots + f_n^{(G)} \times x_n$ sera voisin de $p_1\, x_1 + \ldots\ldots + p_n\, x_n$, c'est-à-dire d'une quantité indépendante de G.

Par exemple, soit y la valeur du point amené en jetant un dé. Cherchons la valeur moyenne V de y. Nous remarquons qu'y peut prendre les valeurs

$$x_1 = 1, \ldots\ldots x_6 = 6.$$

La probabilité pour qu'il prenne chacune de ces valeur est $\frac{1}{6}$. Donc

$$V = \frac{1}{6} \times 1 + \ldots + \frac{1}{6} \times 6 = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2}$$

La valeur moyenne est donc : 3,5. Si on avait cherché la moyenne arithmétique d'un nombre fini d'épreuves, elle n'aurait sans doute pas été la même.

En résumé la valeur moyenne V d'un nombre aléatoire dans une catégorie infinie d'épreuves est égale à la somme des produits de chacune des valeurs distinctes que peut prendre ce nombre aléatoire par la probabilité correspondante. D'où la formule simplifiée :

(16 *bis*) $$V = p_1\, x_1 + \ldots + p_n\, x_n$$

On remarquera que la valeur moyenne de x est la somme des espérances mathématiques de gagner des sommes x_1, x_2, ...

De même pour *la valeur probable*. Pour un groupe G, la médiane s'obtient de la façon suivante : on considère les fréquences $f_1^{(G)}$, $f_2^{(G)}$, ... $f_k^{(G)}$, ... $f_n^{(G)}$, puis les sommes successives des premières fréquences, $f_1^{(G)}$; $f_1 + f_2^{(G)}$; ..... Elles sont de plus en plus grandes. La dernière est égale à 1. Donc, en en prenant un assez grand nombre, on dépassera la valeur $\frac{1}{2}$; soit $f_1^{(G)} + \ldots + f_k^{(G)}$ la première qui dépasse $\frac{1}{2}$. Alors x_k est la valeur médiane. — Appelons m_G cette médiane qui dépend du groupe G. On a par hypothèse :

$$f_1^{(G)} + \ldots + f_{k-1}^{(G)} < f_k^{(G)} + \ldots + f_n^{(G)}$$
$$f_1^{(G)} + \ldots + f_k^{(G)} > f_{k+1}^{(G)} + \ldots + f_n^{(G)}.$$

Lorsque G sera très nombreux, les fréquences seront voisines des probabilités respectives. On aura donc :

$$p_1 + \ldots + p_{k-1} \leqq p_k + \ldots + p_n$$
$$p_1 + \ldots + p_k \geqq p_{k+1} + \ldots + p_n.$$

Donc la médiane (nous dirons : la valeur probable), d'un groupe très nombreux sera égale à la valeur x_k telle qu'il soit aussi probable que X soit $\leqq$ ou $\geqq x_k$.

Pour *les écarts*, on raisonnera de même. Soit a un nombre quelconque, et y un nombre aléatoire : on appellera écart moyen de y avec a, dans une certaine

catégorie d'épreuves, la valeur moyenne du nombre aléatoire $|y - a|$ [1] dans cette catégorie d'épreuves. La limite vers laquelle tend l'écart moyen de y avec a dans un groupe d'épreuves dont le nombre croît indéfiniment sera :

$$p_1 | x_1 - a | + \ldots + p_n | x_n - a | .$$

On démontre, exactement comme dans le cas des groupes d'épreuves (en nombre fini) : 1° que la probabilité pour que l'écart de la valeur aléatoire de y avec un nombre a soit au moins égal à t fois l'écart moyen de y avec a, est au plus égale à $\frac{1}{t}$; 2° que la valeur probable des y est une valeur de a qui rend minimum l'écart moyen de y à partis de a.

L'écart quadratique moyen E de y avec a sera la racine carrée de la valeur moyenne du carré de l'écart $y - a$; c'est donc :

$$\sqrt{p_1 (x_1 - a)^2 + \ldots + p_n (x_n - a)^2}$$

On voit qu'on en obtient une mesure approchée par l'écart quadratique moyen de y avec a dans un groupe d'épreuves de plus en plus nombreux pris au hasard dans la catégorie.

De même que dans le cas d'un groupe d'épreuves fini, on définit l'écart quadratique moyen tout court comme la racine carrée μ de la valeur moyenne du

1. Bien entendu il s'agit de la valeur absolue, abstraction faite du signe. Voir p. 100 note 1.

carré de l'écart du nombre aléatoire y avec sa valeur moyenne V. D'où :

$$\mu^2 = p_1 (x_1 - V)^2 + \ldots + p_n (x_n - V)^2. \tag{18}$$

Lemme de Bienaymé. — La probabilité que le nombre aléatoire y prenne une valeur qui diffère d'un nombre donné a de plus d'un multiple $t\mu$ de l'écart quadratique moyen μ de y et de a est inférieure à $\frac{1}{t^2}$.

La démonstration est la même que dans le cas d'un nombre fini d'épreuves. Rappelons que le théorème ne donne un renseignement utile que si $t > 1$, puisqu'une probabilité est toujours ≤ 1.

Application. — L'écart quadratique moyen avec 6 des points amenés au jeu de dés est :

$$\sqrt{\frac{1}{6}(6-1)^2 + \ldots + \frac{1}{6}(6-6)^2} = \sqrt{\frac{1}{6}(5^2 + \ldots + 1^2)}$$
$$= \sqrt{\frac{1}{6}(1+4+9+16+25)} = \sqrt{\frac{1}{6} \cdot 55} = \sqrt{9,\ldots} = 3,\ldots$$

La probabilité d'un écart avec 6 supérieur ou égal à $t \times 3, \ldots..$ est inférieure ou égale à $\frac{1}{t^2}$. Si nous prenons :

$$t \times 3,\ldots = 5, \quad \frac{1}{t^2} = \left(\frac{3,\ldots}{5}\right)^2 = \frac{55}{6 \times 25} = 0,36\ldots$$

Ainsi, en général, il y aura 36 jets sur 100 *au plus* où le nombre des points obtenus sera 1 (écart avec 6 au

moins égal à 5). En fait nous savons qu'il y en a environ 16 sur 100 $\left(\text{ou } \frac{1}{6}\right)$. Le lemme de Bienaymé ne donne qu'une limite supérieure ; mais alors que le calcul direct, étant facile dans ce cas particulier, donne un résultat plus précis, dans le cas général, où le calcul direct est difficile, on tire de ce lemme une indication déjà très importante.

Comme, encore, dans le cas d'un groupe d'épreuves fini, on voit que le carré de l'écart quadratique moyen E de y avec a est la somme du carré de l'écart quadratique moyen μ à partir de la valeur moyenne V de y et du carré de $(V - a)$:

$$\begin{aligned} & p_1 (x_1 - a)^2 + \ldots + p_n (x_n - a)^2 \\ = & p_1 (x_1 - V)^2 + \ldots + p_n (x_n - V)^2 + (a - V)^2 \end{aligned}$$

ou :

$$E^2 = \mu^2 + (a - V)^2 \tag{17}$$

et que par conséquent le minimum de l'écart quadratique moyen de y à partir d'une valeur variable a a lieu quand a est égal à la valeur moyenne de y.

Dans le jeu de dé, par exemple, l'écart quadratique moyen du nombre de points amenés est :

$$\mu = \sqrt{\frac{1}{6} [1 - v]^2 + \ldots + \frac{1}{6} [6 - v]^2}$$

où v est la valeur moyenne déjà calculée page 148 :

$$v = \frac{7}{2}.$$

Comme v est fractionnaire ($v = 3,5$) il est un peu plus facile de calculer d'abord l'écart quadratique moyen avec une valeur plus simple approchée, 3 par exemple, et d'écrire d'après la formule (17) de la page 152 :

$$\frac{1}{6}[1-3]^2+\ldots+\frac{1}{6}[6-3]^2=\mu^2+(v-3)^2$$

$$\frac{1}{6}(2^2+1^2+1^2+2^2+3^2)=\mu^2+\left(\frac{7}{2}-3\right)^2$$

$$\frac{1}{6}(4+1+1+4+9)=\mu^2+\left(\frac{1}{2}\right)^2$$

$$\mu^2=\frac{1}{6}\,19-\frac{1}{4}=\frac{38-3}{12}=\frac{35}{12}=2,9166$$

$$\mu=\sqrt{2,9166}=1,71.$$

C'est un procédé qu'on pourra souvent employer. Le nombre V est toujours contenu dans l'un des intervalles x_1, x_2, ... x_n, par exemple dans x_k, x_{k+1}. Supposons qu'il soit, par exemple, plus approché de x_k. Alors on écrira :

$$p_1(x_1-x_k)^2+\ldots=(x_k-\mathrm{V})^2+p_1(x_1-\mathrm{V})^2+\ldots$$

En général, dans les statistiques, on a pris x_1, x_2 ..... simples et à intervalles réguliers ; les nombres $(x_1-x_k)^2$, ..., seront donc en général plus faciles à calculer que $(x_1-\mathrm{V})^2$, ... Et la correction $(x_k-\mathrm{V})^2$ sera petite, puisque $(\mathrm{V}-x_k)$ est plus petit que les petits intervalles $|\,x_1-x_2\,|$, $|\,x_2-x_3\,|$, ... et au moins que $|\,x_k-x_{k-1}\,|$.

Puisque l'écart quadratique moyen est minimum à partir de la valeur moyenne, le lemme de Bienaymé donnera un meilleur résultat à partir de cette valeur. Ici la probabilité P d'un écart au moins égal à 5 à partir de 3,5 sera inférieure à $\frac{1}{t^2}$, t étant tel que t fois l'écart quadratique moyen 1,71 soit égal à 5 ; donc

$$t \cdot 1{,}71 = 5 ; \frac{1}{t^2} = \left(\frac{1{,}71}{5}\right)^2 = \frac{2{,}9167}{25} = \frac{35}{300} = 0{,}116.$$

Cette probabilité est plus petite que la valeur trouvée : $\frac{36}{100}$ pour la probabilité d'un écart au moins égal encore à 5, mais à partir de 6. Seulement, ici encore, le calcul direct fournit un résultat encore plus précis que le lemme de Bienaymé ; car un écart $\geqq 5$ à partir de 3,5 pour une valeur qui est égale à 1, 2, ... 6 est impossible. Donc P = 0.

Ici comme précédemment, nous avons choisi un cas où on peut montrer qu'un calcul direct donnerait un résultat plus précis que le lemme de Bienaymé ; mais ceci ne pourra qu'accroître notre confiance. Dans le cas où le calcul direct est trop difficile et où le lemme de Bienaymé nous donnera une petite probabilité, nous saurons alors que le calcul direct, s'il était possible, nous donnerait une probabilité extrêmement petite.

Sommes, produits, carrés, ... de nombres aléa-

toires. — Considérons deux nombres aléatoires X, Y, qui prennent chacun une valeur déterminée dans chaque épreuve d'une catégorie (par exemple ce seront dans chaque jet de dé les points marqués sur les faces supérieure et inférieure). — Supposons qu'on veuille calculer la valeur moyenne d'une certaine fonction de X et Y ; par exemple, celle de leur somme, X + Y, de leur produit, XY, etc. La première quantité X est supposée ne prendre que n valeurs distinctes $x_1 < x_2 \ldots\ldots < x_n$, et Y que r valeurs distinctes, $y_1 < \ldots\ldots < y_r$. La somme X + Y ne pourra donc prendre que les valeurs suivantes : $x_1 + y_1$, $x_1 + y_2$, ... $x_1 + y_r$, $x_2 + y_1$, ... $x_n + y_r$, lesquelles ne sont d'ailleurs pas nécessairement distinctes ; de plus certains de ces couples de valeurs, x_k, y_h, ne seront peut-être jamais pris, si les événements consistant en ce que $X = x_k$ et en ce que $Y = y_h$ ne peuvent se présenter à la fois dans la même épreuve. Ainsi, dans le cas du jeu de dé, quand $X = 1$, Y ne peut pas être égal à 1. En résumé le système des valeurs aléatoires X, Y est tel que X et Y y prennent simultanément telles valeurs, x_k, y_h par exemple, avec une probabilité p_{kh}.

Si le couple x_k, y_h ne peut se présenter, on aura $p_{kh} = 0$. Si on connaît les p_{kh}, on pourra calculer directement la valeur moyenne de toute fonction de X et de Y.

Mais peut-on la calculer, connaissant seulement les probabilités p_1, p_2, ... de x_1, x_2, ... et p'_1, p'_2, ... de y_1, y_2 ..... ?

Considérons la somme $X + Y$. Sa valeur moyenne est :

$$\Sigma\, p_{kh}\,(x_k + y_h)^1,$$

c'est-à-dire :

$$p_{11}(x_1 + y_1) + p_{12}(x_1 + y_2) + \ldots + p_{21}(x_2 + y_1) + p_{22}(x_2 + y_2) + \ldots =$$
$$x_1(p_{11} + p_{12} + \ldots) + x_2(p_{21} + p_{22} + \ldots) + y_1(p_{11} + p_{21} + \ldots) + y_2(p_{12} + p_{22} + \ldots) + \ldots$$

Or $p_{11} + p_{12} + \ldots\ldots$ c'est la probabilité pour que X prenne la valeur x_1 et que Y prenne l'une des valeurs $y_1, y_2, \ldots\ldots$ (cas qui s'excluent mutuellement par hypothèse et où par suite on peut appliquer le théorème des probabilités totales). Donc comme Y prend nécessairement par hypothèse toutes ces valeurs, $p_{11} + p_{12} + \ldots\ldots$ c'est aussi la probabilité pour que X prenne la valeur x_1, indépendamment de la valeur que peut prendre Y. C'est donc p_1. De même $p_{21} + p_{22} + \ldots\ldots = p_2$; $p_{11} + p_{21} + \ldots\ldots = p'_1$ (p'_1 étant la probabilité pour que Y prenne la valeur y_1).

Donc *la valeur moyenne de la somme de deux nombres aléatoires* X, Y, *dans une certaine catégorie d'épreuves, est la somme des valeurs moyennes de chacun de ces deux nombres dans la même catégorie d'épreuves*. Le raisonnement s'étend évidemment à une somme de plus

1. Bien entendu il faut distinguer p_{kh} défini plus haut de la probabilité pour que $X + Y$ ait une valeur égale à $x_k + y_h$.

de 2 nombres. De sorte que si, dans une certaine catégorie d'épreuves, chaque épreuve fournit n nombres aléatoires X_1 X_2, ..., X_n, la somme de leurs valeurs moyennes est égale à la valeur moyenne de leur somme dans cette même catégorie d'épreuves.

Par exemple la production française moyenne par hectare en céréales est égale à la somme des productions moyennes en blé, seigle, etc. Ce résultat a l'air évident. Cependant il ne s'appliquerait pas à la production probable (médiane) par hectare. Supposons par exemple que 2 millions d'hectares produisent 20 hectolitres de blé et 5 hectolitres d'avoine à l'hectare, que 2 autres millions d'hectares produisent 5 hectolitres de blé et 10 hectolitres d'avoine à l'hectare, et que 2 derniers millions d'hectares ne soient point plantés en céréales. En considérant la production d'un hectare en céréales comme une épreuve, on voit que dans l'ensemble des 6 millions d'hectares la valeur probable de la production du blé à l'hectare serait 5 hectolitres, celle de l'avoine, 5 hectolitres, celle du blé et de l'avoine, 15 hectolitres, ce qui n'est pas la somme de $5 + 5$.

Passons au *produit de deux nombres aléatoires* XY ; ce produit prend les valeurs $x_1\, y_1$, $x_1\, y_2$, ... $x_2\, y_1$, ... $x_2\, y_r$, ... $x_n\, y_r$ *avec les mêmes probabilités que précédemment* p_{11}, p_{12}, ... Donc sa valeur moyenne est :

$$(19) \quad p_{11}\, x_1\, y_1 + p_{12}\, x_1\, y_2 + \ldots + p_{21}\, x_2\, y_1 \ldots + p_{nr}\, x_n\, y_r.$$

D'autre part le produit des valeurs moyennes de X et de Y est :

$$
\begin{aligned}
(20) \qquad & (p_1 x_1 + \ldots)(p'_1 y_1 + \ldots) \\
& = p_1 x_1 (p'_1 y_1 + \ldots) + p_2 x_2 (p'_1 y_1 + \ldots) + \ldots \\
& = p_1 p'_1 x_1 y_1 + p_1 p'_2 x_1 y_2 + \ldots + p_n p'_r x_n y_r
\end{aligned}
$$

On voit tout de suite que les deux valeurs (19) et (20) seraient égales si :

$$p_{11} = p_1 p'_1 , \; p_{12} = p_1 p'_2 , \ldots \; p_{kh} = p_k p'_h \ldots$$

autrement dit si la probabilité composée de l'événement consistant en ce que dans une épreuve $X = x_k$, $Y = y_h$, était égale au produit de la probabilité de l'événement consistant en ce que dans une épreuve X prend la valeur x_k, multipliée par la probabilité de l'événement consistant en ce que dans une épreuve Y prend la valeur y_h, (et ceci, quels que soient k et h,) autrement dit si le théorème des probabilités composées est applicable à $p_k \, p'_h$.

Or, on sait qu'il est applicable si les deux derniers événements sont indépendants, mais qu'il n'est pas toujours exact dans le cas général. Si deux nombres aléatoires X, Y, sont indépendants dans une certaine catégorie d'épreuves, la probabilité pour X de prendre une valeur déterminée est indépendante de la valeur de Y. Nous avons alors le théorème suivant :

La valeur moyenne du produit de deux nombres aléatoires INDÉPENDANTS *dans une certaine catégorie d'épreuves est égale au produit des valeurs moyennes de ces deux*

nombres dans cette même catégorie D'ÉPREUVES. — Le théorème ne serait pas exact en général s'il s'agissait de valeurs probables.

Naturellement, les théorèmes énoncés plus haut s'étendraient immédiatement à la somme et au produit de plus de deux nombres aléatoires ; ils s'étendent aussi à la différence de deux nombres aléatoires : sa valeur moyenne est égale à la différence de leurs valeurs moyennes dans la même catégorie d'épreuves. En particulier la valeur moyenne de la différence $X - v$ entre un nombre aléatoire X et sa valeur moyenne v est égale à $v - v$, c'est-à-dire est nulle.

Ceci permet de démontrer une formule très importante.

Soit $X_1, X_2, \ldots X_r$, r nombres aléatoires dont l'ensemble est déterminé pour chaque épreuve dans une certaine catégorie d'épreuves, X leur somme, μ_1, μ_2, ... μ_r et μ les écarts quadratiques moyens respectifs de $X_1, \ldots X_r$ et X. Cherchons à déterminer μ^2. — Appelons v, $v_1, \ldots v_r$ les valeurs moyennes de X, X_1, ... X_r. On a vu (p. 156), pour $r = 2$, — et le raisonnement s'applique évidemment quel que soit r, que : $v = v_1 + \ldots + v_r$ en d'autres termes, la moyenne d'une somme est égale à la somme des moyennes de ses termes dans la même catégorie d'épreuves.

Comme, d'autre part,

$$X = X_1 + \ldots + X_r \quad \text{et} \quad v = v_1 + \ldots + v_r,$$

pour chaque épreuve particulière,

$$X - v = (X_1 - v_1) + \ldots + (X_r - v_r).$$

Supposons d'abord $r = 2$.

$$(X - v)^2 = (X_1 - v_1)^2 + (X_2 - v_2)^2 + 2\,(X_1 - v_1)\,(X_2 - v_2).$$

Par suite la valeur moyenne μ^2 de $(X - v)^2$ est égale à la somme de la valeur moyenne de $(X_1 - v_1)^2$, de la valeur moyenne de $(X_2 - v_2)^2$, soit $\mu_1^2 + \mu_2^2$, et de la valeur moyenne du produit $2\,(X_1 - v_1)\,(X_2 - v_2)$. Si X_1 et X_2 sont indépendants, v_1, v_2 n'étant pas aléatoires puisqu'étant calculés pour la catégorie indéfinie, $X_1 - v_1$, $X_1 - v_2$ sont indépendants, et par suite la valeur moyenne de $(X_1 - v_1)\,(X_2 - v_2)$ est le produit des valeurs moyennes de $X_1 - v_1$ et $X_2 - v_2$ lesquelles sont nulles. Donc $\mu^2 = \mu_1^2 + \mu_2^2$. En opérant ainsi de proche en proche pour $r = 2, 3, \ldots\ldots$, on voit dans le cas général que *le carré de l'écart quadratique moyen d'une somme de nombres aléatoires* INDÉPENDANTS *est égal à la somme des carrés des écarts quadratiques moyens de chacun des termes de cette somme.* En d'autres termes si X_1, X_2, ... X_r sont des nombres aléatoires *indépendants* les uns des autres, dans une même catégorie d'épreuves, et si μ_1, μ_2, ... sont leurs écarts quadratiques moyens, et μ celui de leur somme, on a :

$$\mu^2 = \mu_1^2 + \mu_2^2 + \ldots + \mu_r^2. \tag{21}$$

Moyennant cette formule, un lemme démontré par

Tchebichef se réduit à une forme développée du lemme de Bienaymé. — Considérons des nombres aléatoires indépendants les uns des autres, X_1, X_2, ... X_r, dont l'ensemble est numériquement déterminé par chaque épreuve : la probabilité pour que leur somme $X = X_1 + ... + X_r$ diffère de la somme $v = v_1 + ... + v_r$ de leurs moyennes arithmétiques de plus de t fois la quantité $\sqrt{\lambda_1^2 - v_1^2 + ... + \lambda_r^2 - v_r^2}$ est inférieure à $\frac{1}{t^2}$ en appelant λ_1^2, ... λ_r^2 les valeurs moyennes respectives des carrés X_1^2 ..... X_r^2 de ces nombres.

Nous venons de reproduire (approximativement) l'énoncé de Tchebichef qui est inutilement compliqué. Il suffit en effet de remarquer que

$$\sqrt{\lambda_1^2 - v_1^2 + \ldots + \lambda_r^2 - v_r^2}$$

est égal à l'écart quadratique moyen de X, et v à la moyenne de X. Nous avons déjà prouvé (p. 156) la seconde assertion ; pour la première, il suffit, d'après (21) d'établir que $\mu_1^2 = \lambda_1^2 - v_1^2$, ... ou

$$\lambda_1^2 = \mu_1^2 + v_1^2, \;\ldots\ldots\; \lambda_r^2 = \mu_r^2 + v_r^2.$$

Or λ_1 est l'écart quadratique moyen de x_1 à partir de la valeur zéro, et $| v_1 |$ est l'écart quadratique moyen de sa moyenne v_1 à partir de la valeur zéro. On retombe alors sur un cas particulier d'une formule (17) plus générale (p. 152).

Remarque. — La formule (21) prend une forme

simple quand chacun des écarts $\mu_1, \mu_2, \ldots\ldots \mu_r$ est égal à un même nombre ε. On a alors $\mu^2 = r\varepsilon^2$,

$$\mu = \varepsilon\sqrt{r}. \tag{21 bis}$$

Si par exemple $X_1, X_2, \ldots\ldots X_r$ sont des valeurs approchées de divers nombres $A_1, A_2, \ldots\ldots A_r$ et si les calculs ou les expériences ont été faits avec le même soin, l'écart quadratique moyen μ_1 des valeurs de X_1, sera égal à l'écart quadratique moyen μ_2 des valeurs de $X_2, \ldots\ldots$ Et chacun de ces écarts peut en un certain sens mesurer ou au moins repérer l'ordre de grandeur de l'erreur à craindre sur la grandeur correspondante. Il résulte alors de la formule précédente que l'erreur à craindre sur la somme de ces r grandeurs sera égale à l'erreur à craindre sur chacune multipliée par la racine carrée du nombre de ces grandeurs.

Ce résultat semble en contradiction avec celui qui est fourni par la théorie des approximations numériques : l'erreur absolue sur une somme est au plus égale à la somme des erreurs absolues sur chaque terme de la somme, et par conséquent à $r\varepsilon$ si celles-ci sont au plus égales à ε.

Cela tient à ce que cette dernière théorie fournit une limite supérieure de toutes les erreurs possibles — même exceptionnelles — sur la somme. Au contraire, la formule 21 *bis* donne une idée de l'ordre de grandeur de l'erreur à craindre *en moyenne* sur la somme. Comme les erreurs faites sur chaque terme ne seront pas *en général* de même sens, l'erreur à craindre $\varepsilon\sqrt{r}$ sera plus

petite que la plus grande erreur possible $r\varepsilon$ (par exemple pour une somme de 16 termes la première est 4ε, la seconde 16ε.) Ici encore le bon sens est d'accord avec le calcul; mais le calcul précise les inductions du bon sens.

Détermination empirique de la valeur moyenne et de l'écart quadratique moyen. — Cherchons à déterminer empiriquement au moyen d'un nombre fini d'épreuves la valeur moyenne V et l'écart quadratique moyen μ d'un nombre aléatoire X.

Chaque épreuve de la catégorie d'épreuves envisagée donne une valeur de X qu'on peut considérer comme une première valeur approchée de V. Pour se faire une idée de l'ordre de grandeur de l'erreur à craindre, il suffit de se souvenir que la probabilité d'une erreur absolue $|X - V|$ supérieure à k fois μ est inférieure à $\frac{1}{k^2}$. Autrement dit il est peu probable que l'erreur commise en prenant X comme valeur approchée de V soit notablement supérieure à μ. On peut donc repérer l'erreur à craindre sur V par la grandeur de μ.

Mais la connaissance du résultat de plusieurs épreuves nous permettra-t-elle de calculer V avec plus de précision? Supposons que N épreuves aient donné pour X les valeurs X_1, X_2, ... X_N.

Si ces épreuves sont indépendantes, X_2, X_3, ... X_N sont les résultats d'épreuves quelconques comme celle qui a donné X_1; pour X_1, pour X_2, ... ou pour X_N, la

valeur moyenne est V et l'écart quadratique moyen est μ. Par conséquent la valeur moyenne de $X_1 + \ldots\ldots + X_N$ est NV et par suite V est la valeur moyenne non seulement de X, mais aussi de

$$\frac{X_1 + X_2 + \ldots + X_N}{N}$$

c'est-à-dire de la moyenne arithmétique v_N des valeurs de X dans un groupe de N épreuves.

Une des valeurs de v_N peut donc servir au même titre qu'une des valeurs de X comme valeur approchée de V ; mais le calcul plus compliqué de v_N va être justifié par une exactitude en général plus grande. Pour le montrer calculons l'écart quadratique moyen de v_N, soit λ_N.

Puisque

$$X_1 + \ldots + X_N = N v_N$$

l'écart quadratique moyen de $X_1 + \ldots\ldots + X_N$ est $N\lambda_N$. Or d'après la formule 21 de la page 160, et puisque $X_1 \ldots\ldots X_N$ sont supposés indépendants, le carré de l'écart quadratique moyen de $X_1 + \ldots\ldots + X_N$, soit $(N\lambda_N)^2$ est égal à la somme des carrés des écarts quadratiques moyens de $X_1 \ldots\ldots X_N$, tous égaux à μ^2. Donc :

$$(N\lambda_N)^2 = N\mu^2 \quad \text{d'où} \quad \lambda_N{}^2 = \frac{\mu^2}{N} \quad \text{et}$$

$$\lambda_N = \frac{\mu}{\sqrt{N}} \tag{22}$$

Ainsi l'écart quadratique moyen de v_N est égal au quotient de celui de X par $\sqrt{N}$. *En prenant comme valeur approchée de la valeur moyenne V du nombre aléatoire X, la moyenne arithmétique v_N de N valeurs observées pour X (au lieu d'une seule valeur de X), on divise l'ordre de grandeur de l'erreur à craindre sur V dans une seule observation, par la racine carrée du nombre des valeurs observées.* En prenant la moyenne de 4 épreuves on réduira en général l'erreur de moitié. En prenant la moyenne de 100 épreuves on la réduira au dixième.

Remarque. — La formule précédente (22) a été établie en supposant essentiellement que les observations X_1, X_2, ... sont indépendantes. Il y a lieu, quand on l'applique, de ne pas perdre de vue cette condition. Par exemple, supposons que le nombre aléatoire X soit la longueur d'une amande tirée au hasard d'un sac d'amandes déterminé. On pourra être tenté, pour former v_N, de prendre au hasard une poignée d'amandes et de calculer la moyenne arithmétique de leurs longueurs X_1, X_2, ... X_N. Or, dans les conditions matérielles de ces N épreuves, X_1, X_2, ... X_N ne sont plus indépendants ; ainsi X_2 est la longueur d'une amande *distincte* de la première. Au contraire ces nombres seraient indépendants si, après avoir mesuré la première amande, on la rejetait dans le sac avant de tirer la seconde pour la mesurer.

C'est pourtant sans prendre ce genre de précautions qu'on applique en général la formule, et il est bon de

voir pourquoi les résultats ne sont pas en général mauvais.

Si, par exemple, il y avait sensiblement autant de chances de tirer une amande qu'une autre, la probabilité de cet événement serait $\frac{1}{A}$, A étant le nombre total des amandes. Alors si par exemple on tire successivement N amandes, la probabilité de choisir la deuxième amande était $\frac{1}{A}$ avant toute opération, $\frac{1}{A-1}$ après la première ; la probabilité de choisir la $N^{\text{ième}}$ était $\frac{1}{A}$ avant toute opération, $\frac{1}{A-N+1}$ après les $N-1$ premières. Si donc A est beaucoup plus grand que N, ces probabilités ont peu varié non seulement absolument mais relativement : X_N est sensiblement indépendant de $X_1, \ldots\ldots X_{N-1}$. Mais si N n'est pas beaucoup petit que A, les probabilités auront sensiblement varié — au moins relativement — et dans ce cas la formule (22) pourra devenir notablement inexacte. Dans ce cas, il y aura lieu de procéder à un calcul direct, qui est possible, mais que nous ne ferons pas ici.

Calcul empirique de μ. — La formule précédente (22) montre nettement la croissance de la précision avec le nombre des observations. Mais pour en tirer tout l'avantage qu'elle comporte, il faudrait pouvoir calculer la valeur même de μ.

Comme μ^2 est la valeur moyenne de $(X - V)^2$, on en obtiendrait une valeur empirique approchée, d'après

ce qui précède, en prenant pour μ^2 la moyenne arithmétique de N valeurs de $(X - V)^2$, soit :

$$\mu^2 \triangleq \frac{(X_1 - V)^2 + \ldots + (X_N - V)^2}{N}\,^{1}$$

Mais V n'est pas connu directement ; il semblerait donc naturel de prendre pour μ^2 la valeur obtenue en remplaçant dans le second membre V par v_N la moyenne arithmétique de $X_1, \ldots\ldots X_N$. Cela est en effet admissible si N est très grand. Dans le cas contraire, c'est certainement inexact.

Si, pour mettre en évidence cette inexactitude, on suppose $N = 1$ dans l'expression

$$\frac{(X_1 - v_N)^2 + \ldots + (X_N - v_N)^2}{N}$$

on voit que v_N se réduisant à X_1, cette expression se réduit à zéro, quantité qui non seulement n'est pas égale à μ^2 mais n'en dépend même pas.

Cependant cherchons à utiliser cette expression, que nous appellerons ρ_N^2 et dans ce but calculons sa valeur moyenne.

Pour cela mettons d'abord ρ_N^2 sous la forme déduite de la formule (15) de la page (117).

$$\rho_N^2 = \frac{(X_1 - V)^2 + \ldots + (X_N - V)^2}{N} - (V - v_N)^2$$

1. Le symbole $\triangleq$ désigne ici une égalité approchée. Par exemple : $\pi \triangleq \frac{22}{7}$.

La valeur moyenne de la fraction du second membre est égale à $\frac{\mu^2 + \ldots\ldots + \mu^2}{N} = \mu^2$; nous venons de montrer que celle de $(V - v_N)^2$ est $\frac{\mu^2}{N}$. Donc la valeur moyenne de ρ_N^2 est $\mu^2 - \frac{\mu^2}{N} = \frac{N-1}{N}\mu^2$ et par suite μ^2 est la valeur moyenne de

$$\frac{N}{N-1}\rho_N^2 = \frac{(X_1 - v_N)^2 + \ldots + (X_N - v_N)^2}{N-1}$$

Finalement, si N observations (indépendantes) du nombre aléatoire X donnent les valeurs $X_1, \ldots\ldots X_N$ et si on en calcule la moyenne arithmétique

$$v_N = \frac{X_1 + \ldots + X_N}{N}$$

on a l'égalité approchée

$$\mu^2 \simeq \frac{(X_1 - v_N)^2 + \ldots + (X_N - v_N)^2}{N-1} \tag{23}$$

où le second membre a pour valeur moyenne le carré μ^2 de l'écart quadratique moyen de X.

Comme nous l'avons observé, si N est grand ρ_N^2 peut aussi être considéré comme valeur approchée de μ^2 ; nous voyons bien maintenant que $\frac{N}{N-1}\rho_N^2$ diffère peu de ρ_N^2. Si par exemple on fait 101 observations, $\frac{N}{N-1} = 1{,}01$ et l'erreur relative commise est de un centième. Mais, si l'on n'a qu'une dizaine d'obser-

vations, l'erreur relative sera de un dixième, ce qui n'est pas négligeable.

Valeur moyenne et écarts des fréquences. — Nous allons appliquer les définitions et les propriétés concernant les valeurs typiques et les écarts d'un nombre aléatoire au cas où ce nombre aléatoire mesure *la fréquence d'un événement fortuit.*

Soit un événement fortuit E de probabilité p dans une certaine catégorie C d'épreuves. Appelons p la *probabilité fondamentale*. Effectuons un nombre n d'épreuves. Pour chacune, E se produit ou non. Sur ces n épreuves il y en a r où il se produit, soit r succès : sa répétition est r. La fréquence de E est égale à $\frac{r}{n}$ dans le groupe de n épreuves; r peut prendre toutes les valeurs 0, 1, 2, ... n : la fréquence est comprise entre 0 et 1.

Introduisons la probabilité pour que la fréquence ait une valeur déterminée. Soit P_r la probabilité pour que, sur les n épreuves, il y en ait r favorables. Les théorèmes généraux sur les probabilités nous permettent de la calculer (voir page 186). Ce n'est pas essentiel pour le moment. — Cherchons la valeur moyenne des fréquences. Qu'est-ce à-dire? Chaque fréquence $f = \frac{r}{n}$ sera un nombre aléatoire obtenu dans un catégorie infinie d'épreuves. Au lieu d'une catégorie infinie d'épreuves isolées, envisageons une catégorie infinie G_n de groupes de n de ces épreuves. Soit f la valeur de la fréquence pour un de ces groupes. Cette fréquence

ne peut prendre que les valeurs $0, \frac{1}{n}, \ldots\ldots \frac{n}{n} = 1$. Appelons $P_0, P_1, \ldots\ldots P_n$ les probabilités correspondantes. — Si on recherche maintenant la valeur moyenne V de f dans la catégorie G_n (qui comprend tous les groupes imaginables de n épreuves de la catégorie C), on trouve, d'après la formule (16 *bis*, page 148) :

$$V = P_0 \times 0 + P_1 \times \frac{1}{n} + P_2 \times \frac{2}{n} + \ldots + P_n \times \frac{n}{n}.$$

Pour trouver V, le moyen le plus naturel semble être de calculer $P_0, P_1, \ldots\ldots$ On pourrait le faire au moyen de la formule (26) de la page 187 où P_r serait remplacé par la valeur de ϖ_r^{n}. Mais on peut y arriver directement, et même d'une manière plus simple. Au lieu de calculer V (moyenne des fréquences), je calcule la moyenne des répétitions : r. Toutes les valeurs de f seront proportionnelles à celles de r, car $f = \frac{r}{n}$, n étant donné. Si V' est la moyenne des répétitions dans n épreuves, $V = \frac{V'}{n}$. Essayons de calculer V', moyenne de n résultats. — Appelons X_1 la répétition de E dans les premières épreuves des séries de n épreuves ; nécessairement $X_1 = 0$ ou 1 ; X_2, la répétition de E dans les secondes épreuves des séries de n épreuves : $X_2 = 0$ ou 1. Il en est de même jusqu'à X_n, répétition de E dans les $n^{\text{èmes}}$ épreuves : $X_n = 0$ ou 1. La répétition de E dans le groupe de n épreuves sera : $r = X_1 + X_2 + \ldots\ldots + X_n$: La moyenne de r, ou V', est la

somme des valeurs moyennes de X_1, X_2, ... respectivement (d'après la règle sur la moyenne d'une somme, voir page 156). La moyenne de X_1 est évidemment la probabilité d'un succès dans cette épreuve $= p$, car X_1 prend la valeur 1 avec la probabilité p quand E a lieu, la valeur 0 avec la probabilité complémentaire $(1 - p)$ quand E n'a pas lieu. Par conséquent $V' = p + p + \ldots\ldots + p = np$. Le même raisonnement peut se présenter d'une manière plus intuitive en faisant intervenir l'espérance mathématique. V' est la moyenne des répétitions dans n épreuves, donc l'espérance mathématique du gain consistant à recevoir r francs s'il y a eu r succès. C'est la somme des espérances mathématiques des gains consistant à recevoir 1 franc dans la 1re épreuve quand il y a succès, plus 1 franc dans la 2e épreuve quand il y a succès, et ainsi de suite. Ces espérances mathématiques sont : $1 \times p$. Donc $V' = np$. — D'où $V = \frac{V'}{n} = p$. — Ainsi *la moyenne V de la fréquence d'un événement fortuit dans* n *épreuves est égale à la probabilité de cet événement fortuit.* — En soi ce résultat ne nous apprend pas grand'chose de plus que la définition de la probabilité ; mais il est précieux pour les calculs numériques à faire. Nous allons en avoir une utilisation immédiate.

Ecart quadratique moyen des fréquences. — Soit μ l'écart quadratique moyen de la fréquence f dans la catégorie G_n de séries de n épreuves. μ^2 est donné par

la formule (18) p. 151 (où ce qui joue le rôle du nombre aléatoire X est actuellement la fréquence f, dont les seules valeurs distinctes ne sont plus $x_1, \ldots x_i$, mais $0, \frac{1}{n}, \ldots \frac{n}{n} = 1$. Nous devons donc y remplacer la moyenne V de X par la moyenne p de f, et remplacer également $p_1, \ldots p_n$ par les probabilités respectives $P_0, P_1, \ldots P_n$ des valeurs des fréquences). D'où :

$$\mu^2 = (0 - p)^2 \times P_0 + \left(\frac{1}{n} - p\right)^2 \times P_1 + \ldots$$
$$+ \left(\frac{m}{n} - p\right)^2 \times P_m + \ldots + (1 - p)^2 \times P_n.$$

Il ne reste plus qu'à remplacer dans cette expression $P_0, P_1, P_2, \ldots$ par leurs valeurs en fonction de n. Mais on peut aussi le calculer plus simplement en se servant de la formule qui donne l'écart quadratique moyen d'une somme (voir p. 160). Dans ce but calculons d'abord l'écart quadratique moyen des répétitions $X = nf$ de l'événement E au cours de n épreuves. Son carré sera $n^2\mu^2$, d'où, en multipliant les deux membres de la formule précédente par n^2

$$n^2 \mu^2 = (pn - 0)^2 P_0 + (pn - 1)^2 P_1 + \ldots + (pn - m)^2 P_m$$
$$+ \ldots (pn - n)^2 P_n$$

Comme nous l'avons vu plus haut, la répétition de l'événement E dans un groupe de n épreuves est égale à la répétition (0 ou 1) de E dans la 1re épreuve, soit X_1, plus la répétition (0 ou 1) de E dans la 2e épreuve, soit X_2, etc. Donc le nombre aléatoire X est la somme

des nombres aléatoires X_1, X_2, ... Le carré de l'écart quadratique moyen de X est, d'après la page 160, la somme des carrés des écarts quadratiques moyens des nombres aléatoires X_1, X_2, etc. Or X_1 par exemple est un nombre aléatoire qui peut prendre la valeur 1 quand E a lieu, avec la probabilité p, et la valeur 0 quand E n'a pas lieu, avec la probabilité $q = 1 - p$. La valeur moyenne de X_1 est p (voir p. 171), le carré de son écart quadratique moyen est donc, d'après la formule 21 de la page 160 : $(1-p)^2 p + (0-p)^2 q = q^2 p + p^2 q = pq(p+q) = pq$. — De même les carrés des écarts quadratiques moyens de X_2, ... X_n sont égaux à pq, et par suite celui de X est npq. On a finalement :

$$n^2 \mu^2 = npq\,; \qquad \mu^2 = \frac{pq}{n}\,;$$

$$\mu = \sqrt{\frac{pq}{n}}\,; \quad \left(\mu = \sqrt{\frac{p(1-p)}{n}}\right). \tag{24}$$

Cette formule est une des plus importantes du Calcul des Probabilités. Elle permet de *calculer l'écart quadratique moyen de la fréquence dans* n *épreuves d'un événement de probabilité connue.*

Cas où p *est petit.* — Nous pouvons remarquer en passant que l'écart quadratique moyen des répétitions qui est $n\mu = \sqrt{npq}$ peut s'écrire aussi $\sqrt{r_0 q}$ en appelant r_0 la valeur moyenne np de la répétition de l'événement favorable dans les n épreuves. Il en résulte que, dans le cas particulier où l'événement est peu probable, on peut évaluer l'écart quadratique moyen des répétitions

sans connaître le nombre n des épreuves lui-même. En effet, dans ce cas, q est voisin de 1, et l'écart se réduit sensiblement à $\sqrt{r_0}$. Pour un groupe déterminé de n épreuves dont r sont favorables, r_0 sera en général voisin de r si n est grand, de sorte qu'approximativement $\mu = \sqrt{r}$.

Ceci est utile dans certaines statistiques. Par exemple on connaît le nombre annuel r des ouvriers qui ont été tués dans un accident du travail sans connaître exactement le nombre total n des ouvriers. Mais on sait que n est considérablement plus grand que r, donc que p, qui est voisin de $\frac{r}{n}$, est petit. Par suite l'écart quadratique moyen du nombre annuel r des ouvriers morts par accidents est environ la racine carrée de ce nombre r. Par exemple, de 1900 à 1907 le nombre r signalé aux inspecteurs du travail a varié entre 1.300 et 1.800, et son écart μ est approximativement compris entre $\sqrt{1.300} = 36,0$ et $\sqrt{1.800} = 42,4$.

Ecart moyen, écart probable. — Le calcul de l'écart moyen σ de f et celui de l'écart probable ρ supposent le calcul intégral. Nous nous en tiendrons à donner, pour mémoire, leurs valeurs, soit : $\sigma = \sqrt{\frac{2}{\pi}} \times \mu = \sqrt{\frac{2pq}{\pi n}}$; $\rho = k\mu$; les valeurs approchées à $\frac{1}{10.000}$ près des coefficients $\sqrt{\frac{2}{\pi}}$ et k étant 0,7981 et 0,6745.

Et, plus approximativement : $\sigma = \frac{4}{5}\,\mu$, $\rho = \frac{2}{3}\,\mu$;

de sorte que $\rho < \sigma < \mu$, ces trois nombres étant sensiblement proportionnels à : $\frac{2}{3}$, $\frac{4}{5}$ et 1, ou encore à : 10, 12 et 15.

Théorème de Bernoulli. — Nous sommes maintenant en mesure de démontrer très simplement le théorème de Bernoulli, sans faire intervenir les intégrales dont on se sert d'habitude. Il nous suffira de combiner le lemme de Bienaymé avec la formule $\mu = \sqrt{\frac{pq}{n}}$ qui donne l'écart quadratique moyen μ de la fréquence $f = \frac{r}{n}$ d'un événement de probabilité p dans un groupe de n épreuves.

La valeur de μ montre déjà en gros comment se produit la répartition des fréquences. μ est un écart typique. Mais qu'arrive-t-il si on compare des séries d'épreuves qui ont des nombres d'épreuves différents ? Si on suppose p donné, q est aussi donné. Si n croît, $\frac{pq}{n}$ décroît, et aussi sa racine : μ décroît. Au contraire $\mu' = \sqrt{npq}$ croît. Quand le nombre n d'épreuves croît, l'écart quadratique moyen des fréquences décroît, et l'écart quadratique moyen du nombre des succès croît.

J'ai une probabilité déterminée, p. Si je prends les différentes valeurs possibles de f, l'écart de l'ensemble de ces valeurs avec la probabilité décroît quand le nombre de ces épreuves croît, — et décroît en tendant

vers zéro. Les fréquences se groupent de plus en plus près de la probabilité. Cela ne veut pas dire qu'elles vont en être toutes voisines, mais que la plus grande partie d'entre elles en seront de plus en plus proches.

Essayons de préciser, en cherchant la probabilité pour que la fréquence diffère de la probabilité d'une quantité inférieure à une quantité donnée (ou de plus d'une quantité donnée). Appelons $\mathfrak{P}_\varepsilon$ la probabilité pour que $f - p$ soit, en valeur absolue, $\geqq$ une quantité donnée ε. Je me sers du lemme de Bienaymé qui me donne une limite supérieure. Si j'appelle θ_t la probabilité que $|f - p| \geqq \mu t$, cette probabilité est inférieure ou égale à $\frac{1}{t^2}$. Si je prends ε tel que $\mu t = \varepsilon$, alors $\mathfrak{P}_\varepsilon = \theta_t$, et puisque $\frac{1}{t} = \frac{\mu}{\varepsilon}$ et $\mu^2 = \frac{pq}{n}$:

$$\mathfrak{P}_\varepsilon \leqq \frac{\mu^2}{\varepsilon^2}, \quad \mathfrak{P}_\varepsilon \leqq \frac{pq}{n\varepsilon^2} .$$

Si n croît indéfiniment, $\mathfrak{P}_\varepsilon$ tend vers zéro. Par conséquent la probabilité $\mathfrak{P}_\varepsilon$ pour que la fréquence diffère de la probabilité de plus d'une quantité donnée tend vers zéro quand on prend un nombre d'épreuves qui croît indéfiniement.

Si, par conséquent, on cherche la probabilité $\mathfrak{P}'$ pour que $|f - p|$ soit $\geqq \frac{1}{100}$ par exemple, on voit que $\mathfrak{P}' \leqq \frac{p.q.10.000}{n}$. Si le nombre des épreuves n'est pas très grand, le second membre n'est pas très petit. Donc $\mathfrak{P}'$ (qui est en tout cas $\leqq 1$) pourra être voisin de 1, la

fréquence f pourra être souvent assez différente de la probabilité p. Mais si n croît, $p.q.$ 10.000 reste fixe, et $\frac{p.q.10.000}{n} \to 0$; on pourra par exemple prendre n assez grand pour que $\frac{p.q.10.000}{n} < \frac{1}{1.000}$, ou $\frac{p.q.10^7}{n} < 1$; il suffira de prendre $n > p.q.10^7$. Si, par exemple, il s'agit de la fréquence de l'événement consistant à tirer un as d'un jeu de 32 cartes, $p = \frac{1}{8}$, $q = \frac{7}{8}$, $n > \frac{7 \times 10^7}{64}$, on peut prendre $n = 10^7$. Inversement la probabilité que la fréquence dans dix millions de jeux et la probabilité $\frac{1}{8}$ diffèrent de plus de $\frac{1}{100}$ est une probabilité plus petite que $\frac{1}{1.000}$ [1].

D'une façon générale, quel que soit l'événement fortuit qui a la probabilité p, si on se donne à l'avance deux nombres ε, α, on peut trouver n de sorte que la probabilité $\mathfrak{P}_\varepsilon$ pour que $|f - p| > \varepsilon$ soit un nombre $\mathfrak{P}_\varepsilon < \alpha$. Il suffira de prendre, $n > \frac{pq}{\alpha\varepsilon^2}$.

Mais on peut traduire encore ce résultat d'une autre façon : donnons-nous ε, α ; aussi petit que soit n, $\mathfrak{P}_\varepsilon$ sera

1. En fait, la formule de Laplace que nous étudierons plus loin permettrait de montrer que cette probabilité de $\frac{1}{1.000}$ s'obtient déjà pour un nombre de parties notablement plus petit, puisqu'elle est atteinte d'après cette formule pour un nombre de parties approximativement égal à 215.

$< \alpha$ si n est suffisamment grand. Donc, ε étant donné, $\mathfrak{P}_\varepsilon$ tend vers zéro quand n croît indéfiniment. Ou encore :

Dans une catégorie d'épreuves où un événement fortuit E *a une probabilité définie* p, *lorsque le nombre d'épreuves grandit indéfiniment, il devient infiniment peu probable que la fréquence dans ce nombre d'épreuves diffère de la probabilité* p *de plus d'une quantité arbitraire fixée d'avance*. C'est l'énoncé du théorème de Bernoulli.

Il est essentiel de bien comprendre la signification de ce théorème. — Supposons que nous fassions abstraction de la définition de la probabilité donnée au début. Soit un événement fortuit E et une catégorie d'épreuves C. Définissons la probabilité p de E dans C. Choisissons cette définition de façon arbitraire, pourvu qu'elle satisfasse aux conditions suivantes : p sera compris entre 0 et 1 ; $p = 0$ quand la réalisation est impossible ; $p = 1$ quand elle est certaine. En outre la définition est supposée telle que les théorèmes des probabilités totale et composée subsistent : ces deux théorèmes deviennent des conditions auxquelles doit satisfaire la définition de la probabilité. Alors on voit, si l'on se reporte aux raisonnements successifs que nous avons employés, qu'on peut encore démontrer le théorème de Bernoulli, savoir que si ε est un nombre positif donné, si $P_n^{(\varepsilon)}$ est la probabilité pour que la fréquence f_n de E dans une série de n épreuves diffère de la probabilité fondamentale p d'au moins ε, alors $P_n^{(\varepsilon)}$ tend vers 0 quand n croît indéfiniment. — Si j'opère ainsi, je ne suppose pas d'autre

rapport entre la définition de la probabilité et de la fréquence que celui-ci : dans le cas où il y a certitude, ou impossibilité, la probabilité et la fréquence sont égales. — Mais si on se borne à cela, on ne peut rien tirer de la théorie au point de vue de la pratique. Je ne sais pas quel rapport physique il y a entre $P_n^{(\varepsilon)}$ et la fréquence F_n^ε des groupes de n épreuves où $| f_n - p | > \varepsilon$.

Supposons au contraire non seulement que cette probabilité est soumise aux conditions, posées *a priori*, que nous avons dites, mais encore qu'elle est définie comme nous l'avons expliqué au début de cet ouvrage, comme une valeur typique autour de laquelle viennent se grouper les fréquences, et vers laquelle tend pratiquement la fréquence d'un événement dans n épreuves quand n croît indéfiniment; alors c'est l'inverse : le théorème de Bernoulli ne nous apprend rien non plus dans ce second cas, parce qu'il se borne à énoncer sous une autre forme ce que nous avons admis à la base comme vérité d'expérience : il répète la loi du hasard.

Plaçons-nous maintenant dans un cas intermédiaire. Abandonnons de nouveau notre définition du début de cet ouvrage, et plaçons-nous dans l'hypothèse précédente. Mais supposons que la probabilité satisfait à une nouvelle condition, que la fréquence de E dans un grand nombre d'épreuves est voisine de la probabilité *quand la probabilité est petite* ; que, lorsque nous disons qu'elle est petite, cela équivaut à affirmer que l'événement se produira très rarement. Alors le théorème de Bernoulli nous apporte un résultat nouveau.

Distinguons, en effet : 1° la probabilité P_N^{ε} que la fréquence f d'un événement s'écarte de plus de ε de la probabilité p de cet événement ; 2° cette probabilité p au sujet de laquelle on pose la question : quand la fréquence f de l'événement s'en écarte-t-elle de plus de ε ? Bernoulli démontre que la première probabilité P_N^{ε} est petite, pour un très grand nombre d'épreuves, c'est-à-dire (d'après la nouvelle condition) que la fréquence correspondante F_N^{ε} est alors également petite. Il en résulte que, pour un très grand nombre d'épreuves encore, il se produira très rarement (fréquence F_N^{ε} petite) que la fréquence f de l'événement s'écarte sensiblement de la probabilité p de cet événement, *même si celle-ci n'est pas petite*. Comme l'a fait remarquer M. Castelnuovo, le théorème de Bernoulli permet de démontrer la loi du hasard dans le cas général, si on l'admet (ou si on la vérifie expérimentalement) dans un certain cas particulier : celui des petites probabilités fondamentales p.

M. Borel imagine, quelque part, que des singes tapent sur le clavier d'une machine à écrire. Quelle est la probabilité pour qu'à la longue on voie, sur les feuilles de papier noircies au hasard, se détacher une phrase déterminée, par exemple l'énoncé d'un théorème d'Euclide ? Certes, cette probabilité est infiniment petite (bien qu'on puisse la calculer cependant). Et notre expérience, d'autre part, nous permet de dire que cet événement est tout à fait invraisemblable, ou, en d'autres termes, que la fréquence de cet événement est du même ordre que

la probabilité correspondante, c'est-à-dire infiniment petite aussi. Ainsi nous savons, par expérience, que la fréquence d'un événement, dans un très grand nombre d'épreuves, est voisine de sa probabilité, quand celle-ci est très petite (quand elle se rapproche de l'impossibilité). — Considérons maintenant une probabilité qui n'est pas très petite, par exemple la probabilité d'amener double six au jeu de dés. Sans doute, si nous jouions un très grand nombre de parties, nous trouverions que la fréquence de l'événement qui consiste à amener double six ne s'écarte pas sensiblement, en moyenne, de cette probabilité (calculée). Pourtant si, dans une partie de dix coups, nous amenons trois ou quatre fois double six (au lieu de l'amener une fois seulement en trente-six coups), nous n'éprouvons pas le même sentiment de stupeur que si les singes dactylographes, dans le temps limité où nous les observons, écrivaient l'énoncé du théorème en question. En effet, lorsqu'une probabilité n'est pas très petite, nous ne nous étonnons pas que l'événement se produise, dans un nombre limité d'épreuves, avec une fréquence qui déborde sensiblement cette probabilité. C'est ici que le théorème de Bernoulli intervient, pour substituer aux évaluations incertaines, fondées sur des expériences peu nombreuses, une détermination précise : il nous permet de conclure du premier rapport entre probabilité et fréquence dans le cas des singes dactylographes (où l'une et l'autre se rejoignent dans l'infiniment petit et équivalent pratiquement à zéro) au second, dans le cas où nous savons que le

double six a une probabilité qui n'est pas très petite, et de prévoir que, dans ce cas également, pour un très grand nombre d'épreuves, l'écart entre probabilité et fréquence sera extrêmement réduit.

EXERCICE

M. Julin a mesuré les longueurs (en dixièmes de millimètre) de deux lots d'amandes tirées au hasard d'un même bocal. Nous avons rangé par ordre de grandeur les longueurs observées dans les deux lots.

Premier lot de 40 amandes :

254, 267, 271, 281, 285, 291, 292, 295, 299, 300, 305, 305, 307, 310, 314, 315, 315, 318, 322, 323, 325, 328, 329, 331, 333, 334, 339, 343, 352, 354, 364, 367, 374, 374, 375, 382, 382, 396, 396, 400.

Deuxième lot de 54 amandes :

258, 262, 265, 267, 269, 276, 291, 294, 295, 295, 299, 299, 299, 302, 304, 307, 308, 310, 310, 310, 312, 313, 314, 318, 321, 324, 325, 325, 325, 326, 329, 329, 329, 330, 331, 333, 334, 335, 339, 344, 348, 349, 352, 356, 359, 359, 362, 362, 363, 364, 365, 366, 369, 372.

1° On remarque qu'entre la plus petite et la plus grande, il y a une différence de 146 (soit $1^{cm},46$).

Pourtant, en calculant la médiane de chaque lot, on obtiendra le même nombre à 2 dixièmes de millimètre près.

2° Calculer empiriquement l'écart quadratique moyen μ de la longueur de ces amandes en appliquant la formule (23) de la page 168 à deux petits groupes choisis *au hasard* dans

l'ensemble des 94 amandes. Par exemple les deux groupes suivants, obtenus par tirage au sort :

291 ; 364 ; 367 ; 400.
254 ; 258 ; 265 ; 281 ; 295 ; 299 ; 323 ; 325 ; 334 ; 335.

On verra que quoique les moyennes arithmétiques v_4 et v_{10} relatives à ces deux groupes diffèrent assez largement, les valeurs approchées de μ qu'ils fournissent sont assez voisines.

CHAPITRE VI

ÉPREUVES RÉPÉTÉES

Exemple. — On choisit au hasard cinq personnes âgées de 40 ans ; quelle est la probabilité qu'il en reste deux en vie au bout d'une année?

Admettons que les cinq personnes soient prises au hasard dans une collectivité où la probabilité, pour une personne de 40 ans, de mourir avant 41 ait une valeur déterminée q, soit par exemple $q = 0,01$; la probabilité de survie sera $p = 1 - q = 0,99$.

Supposons que les initiales des noms des cinq personnes soient par exemple A, B, C, D, E et cherchons pour commencer la probabilité pour qu'au bout d'un an A et B seules survivent. Alors cette probabilité sera, d'après le théorème des probabilités composées, égale au produit de deux facteurs égaux à la probabilité 0,99 de survie de A et de B séparément, par trois facteurs égaux à la probabilité 0,01 de décès de C, D, E séparément, — ce sera donc : $(0,99)^2 \times (0,01)^3$.

Mais on obtiendrait évidemment la même valeur pour la probabilité qu'au bout d'un an survivent seuls C, D ou celle que survivent seuls B et E, etc.

Or, la probabilité cherchée est évidemment la somme de toutes ces probabilités, c'est-à-dire qu'elle est égale à autant de fois $(0,99)^2 \times (0,01)^3$ qu'on peut former de groupes de deux des cinq personnes A B C D E,

AB, AC, AD, AE ; BC, BD, BE; CD, CE; DE

c'est-à-dire dix fois.

Finalement la probabilité cherchée, qu'on peut désigner par $\varpi_2^{(5)}$ a pour valeur

$$\varpi_2^{(5)} = (0,99)^2 \times (0,01)^3 \times 10 = 0,0000098.$$

Pour la solution nous avons compté directement le nombre de groupes de deux des cinq personnes. Si nous désignons suivant la notation générale de la page 42, ce nombre par C_5^2, nous voyons qu'on a

$$\varpi_2^{(5)} = p^2 q^3 C_5^2.$$

Nous allons généraliser cette formule. Mais auparavant remarquons bien que, pour que le raisonnement soit correct, il faut supposer que la probabilité de décès pour chacune des personnes est indépendante du sort de chacune des autres. Les cinq personnes devront être bien choisies au hasard ; si au contraire on avait soin de prendre cinq personnes en relations quotidiennes, comme par exemple les membres d'une même famille, le risque de contagion aurait pour effet d'augmenter la probabilité de décès d'une personne en cas de décès de l'autre.

Problème général. — Cherchons la probabilité $\varpi_r^{(n)}$ pour qu'un même événement fortuit E se répète r fois

au cours de n épreuves, connaissant seulement la probabilité p de E.

Il y a lieu d'ailleurs de préciser : p est la probabilité de E dans une certaine catégorie C d'épreuves. A chaque épreuve de C, E se produit une fois ou ne se produit pas. $\varpi_r^{(n)}$ est la probabilité que sur n épreuves de C, E se produise dans r épreuves ; de sorte que $\varpi_r^{(n)}$ est mesurée dans une catégorie C_n d'épreuves dont chacune est constituée par un groupe de n épreuves de C.

Ainsi dans l'exemple précédent, une épreuve de C consiste à choisir une personne de 40 ans au hasard dans la collectivité considérée. Une épreuve de C_n, ici C_5, consiste à choisir au hasard un groupe de cinq personnes de 40 ans dans cette collectivité.

Cherchons d'abord la probabilité pour que l'événement fortuit se présente dans certaines épreuves spécifiées et en nombre r, et par suite que E ne se présente pas dans les $n - r$ autres. Alors la probabilité intermédiaire cherchée est, d'après le théorème des probabilités composées, égale au produit de r facteurs égaux à p et de $n - r$ facteurs égaux à $q = 1 - p$; c'est donc $p^r q^{n-r}$.

La probabilité $\varpi_r^{(n)}$ est la somme des probabilités précédentes relatives aux différents groupes de r épreuves distinctes qu'on peut former au moyen des n épreuves considérées. Donc $\varpi_r^{(n)}$ est égal à autant de fois $p^r q^{n-r}$ qu'il y a de tels groupes ; nous avons (page 42) désigné ce nombre de groupes par la notation C_n^r. On a donc :

$$\varpi_r^{(n)} = C_n^r \, p^r \, q^{n-r}. \tag{25}$$

Nous savons (page 42) que C_n^r est le quotient du produit de r entiers décroissants à partir de n par le produit de r entiers croissants à partir de l'unité :

$$(1)' \qquad C_n^r = \frac{n(n-1)\ldots(n-r+1)}{1.2.3\ldots(r-1)r}.$$

En sorte, qu'on a l'expression développée de $\varpi_r^{(n)}$:

$$(26) \qquad \varpi_r^{(n)} = \frac{n(n-1)\ldots(n-r+1)}{1.2\ldots(r-1)r} p^r q^{n-r}.$$

Il faut remarquer que pour $r = n$, C_n^n se réduit à l'unité. Pour $r = 0$, l'expression (1′) de C_n^r n'a plus de sens, mais il est directement visible que, $\varpi_0^{(n)}$ étant la probabilité pour que dans n épreuves l'événement contraire à E se présente seul à chaque épreuve, on a $\varpi_0^{(n)} = q^n$, de sorte qu'il suffira de convenir d'attribuer à C_n^0 la valeur 1 pour que la formule (25) soit valable dans tous les cas [1].

Applications. I. — Quelles sont les probabilités $\varpi_0^{(6)}$, $\varpi_1^{(6)}$, $\varpi_2^{(6)}$, $\varpi_3^{(6)}$, $\varpi_4^{(6)}$, $\varpi_5^{(6)}$, $\varpi_6^{(6)}$ pour que, sur six personnes de 40 ans, survivent au bout d'une année : 0 ; 1 ; 2 ; 3 ; 4 ; 5 ; ou respectivement 6 d'entre elles, sachant que le taux annuel de mortalité à 40 ans est égal à 1 p. 100 dans la collectivité considérée.

On a ici $q = 0{,}01$; $p = 1 - q = 0{,}99$. Il suffit

1. Tout lecteur connaissant le développement du binome de Newton pourra maintenant adopter la règle mnémotechnique suivante : les probabilités pour qu'un événement E de probabilité p se répète 0, 1,..... r,..... n fois dans n épreuves sont les coefficients correspondants dans le développement du binome $(qx + p)^n$, (où $q = 1 - p$), suivant les puissances de x.

d'appliquer la formule (26) avec $n = 6$, $r = 0; 1; \ldots; 6$. On a ainsi :

$$\varpi_0^{(6)} = q^6 ; \varpi_1^{(6)} = \frac{6}{1} q^5 p ; \varpi_2^{(6)} = \frac{6 \times 5}{1 \times 2} q^4 p^2 ;$$

$$\varpi_3^{(6)} = \frac{6 \times 5 \times 4}{1 \times 2 \times 3} q^3 p^3 ; \varpi_4^{(6)} = \frac{6 \times 5 \times 4 \times 3}{1 \times 2 \times 3 \times 4} q^2 p^4 ;$$

$$\varpi_5^{(6)} = \frac{6 \times 5 \times 4 \times 3 \times 2}{1 \times 2 \times 3 \times 4 \times 5} q p^5 ; \varpi_6^{(6)} = p^6$$

ou en remplaçant q par 0,01, p par 0,99, et simplifiant :

$$\varpi_0^{(6)} = (0{,}01)^6 ; \varpi_1^{(6)} = 6 \times (0{,}01)^5 \times 0{,}99 ;$$
$$\varpi_2^{(6)} = 15 \times (0{,}01)^4 (0{,}99)^2 ; \varpi_3^{(6)} = 20 \times (0{,}01)^3 \times (0{,}99)^3 ;$$
$$\varpi_4^{(6)} = 15 \times (0{,}01)^2 \times (0{,}99)^4 ; \varpi_5^{(6)} = 6 \times 0{,}01 \times (0{,}99)^5 ;$$
$$\varpi_6^{(6)} = (0{,}99)^6 .$$

Sans qu'il soit besoin de faire le calcul, on voit sur ces expressions que la probabilité pour que toutes ces six personnes soient encore en vie au bout de l'année est plus grande que toutes les autres.

II. — Quelles sont les probabilités pour qu'au cours d'un groupe de cinq jets de dés on amène 0, 1,..... 5 fois le point six.

Si le dé est bien construit, on pourra admettre que la probabilité d'amener le point six en jetant un dé est égale à $\frac{1}{6}$.

On a donc ici, $p = \frac{1}{6}$, $q = 1 - p = \frac{5}{6}$; $n = 5$.

Donc, en appliquant encore la formule (26) :

$$\varpi_0^{(5)} = \left(\frac{5}{6}\right)^5 = 0{,}402 ; \varpi_1^{(5)} = \frac{5}{1} \left(\frac{5}{6}\right)^4 \frac{1}{6} = 0{,}402 ;$$

$$\varpi_2^{(5)} = \frac{5 \times 4}{1 \times 2}\left(\frac{5}{6}\right)^3 \left(\frac{1}{6}\right)^2 = 0{,}161 ; \varpi_3^{(5)} = \frac{5 \times 4 \times 3}{1 \times 2 \times 3}\left(\frac{5}{6}\right)^2 \left(\frac{1}{6}\right)^3$$

$$= 0{,}032 ; \varpi_4^{(5)} = \frac{5 \times 4 \times 3 \times 2}{1 \times 2 \times 3 \times 4}\,\frac{5}{6}\left(\frac{1}{6}\right)^4 = 0{,}0032 ; \varpi_4^{(5)} = 0{,}00013.$$

On voit qu'il y a autant de chances de n'amener jamais ou de n'amener qu'une fois le point six au cours de cinq jets, et qu'il est assez peu probable d'amener plus d'une fois le point six.

Comparaison des probabilités des diverses répétitions. — Nous avons défini la fréquence f d'un événement dans un groupe de n épreuves comme le quotient par n de la répétition r de cet événement dans le groupe : $f = \frac{r}{n}$. Et nous avons admis comme un fait d'expérience. sous le nom de loi du hasard, que la fréquence f d'un événement fortuit dans un groupe très nombreux d'épreuves est généralement voisine de la probabilité p de cet événement. Autrement dit il est improbable que f ne soit pas voisin de p, ou encore la probabilité $\varpi_r^{(n)}$ pour que la répétition r corresponde à une fréquence $\frac{r}{n}$ éloignée de p doit être petite.

Ces remarques nous placent maintenant devant le problème suivant : vérifier que les expressions trouvées pour $\varpi_r^{(n)}$ s'accordent avec cette prévision ; préciser celle-ci. Grâce aux valeurs précises des $\varpi_r^{(n)}$, il nous sera possible d'observer ce qui se passe même dans les cas où n n'est pas très grand.

Nous sommes donc amenés à comparer les valeurs de $\varpi_0^{(n)}$, $\varpi_1^{(n)}$,... $\varpi_r^{(n)}$,..... c'est-à-dire à étudier la variation de $\varpi_r^{(n)}$ avec r.

Variation de $\varpi_r^{(n)}$ avec r. — Dans le cas d'un groupe peu nombreux, le moyen le plus naturel d'étudier la variation de $\varpi_r^{(n)}$ avec r est de calculer effectivement $\varpi_0^{(n)}$, $\varpi_1^{(n)}$,..... Nous voyons ainsi immédiatement la réponse à donner dans les deux exemples qu'on vient d'étudier.

Mais, si n est grand, le calcul direct devient extrêmement long ou même impraticable. Si, par exemple, $n = 10.000$, valeur qui n'a rien d'excessif dans les applications, il faudrait calculer chacune des 10.001 valeurs $\varpi_0^{(10.000)}$, $\varpi_1^{(10.000)}$,..... $\varpi_{10\,000}^{(10\,000)}$. Il y aura donc avantage à essayer d'un moyen détourné et qui conduise, si possible, à un résultat général, ce que la loi du hasard nous permet d'espérer.

Comme les diverses valeurs de $\varpi_r^{(n)}$ sont des fractions dont les termes sont des produits d'un nombre éventuellement considérable de facteurs, il est naturel, pour déterminer la plus grande de deux valeurs de $\varpi_r^{(n)}$, de calculer non pas leur différence, mais leur quotient, et de voir si celui-ci est supérieur ou inférieur à l'unité.

Formons donc le rapport de deux nombres successifs $\varpi_{r+1}^{(n)}$ et $\varpi_r^{(n)}$. Si, dans (26), on remplace r par $r+1$, on a :

$$\varpi_{r+1}^{(n)} = \frac{n\,(n-1)\ldots(n-r)}{1.2\ldots(r+1)}\,p^{r+1}\,q^{n-r-1}. \tag{27}$$

En divisant les valeurs (27) et (26) l'une par l'autre et simplifiant, on a donc :

$$\frac{\varpi_{r+1}^{(n)}}{\varpi_r^{(n)}} = \frac{n-r}{r+1}\frac{p}{q}. \tag{28}$$

On aurait de même

$$\frac{\varpi_r^{(n)}}{\varpi_{r-1}^{(n)}} = \frac{n-r+1}{r}\frac{p}{q}. \tag{29}$$

On voit que, lorsque r croît, $\frac{n-r}{r+1}$ décroît, donc aussi $\frac{\varpi_{r+1}^{(n)}}{\varpi_r^{(n)}}$; autrement dit :

$$\frac{\varpi_1^{(n)}}{\varpi_0^{(n)}} > \frac{\varpi_2^{(n)}}{\varpi_1^{(n)}} > \ldots > \frac{\varpi_{r+1}^{(n)}}{\varpi_r^{(n)}} > \ldots > \frac{\varpi_n^{(n)}}{\varpi_{n-1}^{(n)}}. \tag{30}$$

Suivant que le rapport $\frac{\varpi_{r+1}^{(n)}}{\varpi_r^{(n)}}$ est supérieur, égal ou inférieur à l'unité, $\varpi_{r+1}^{(n)}$ sera supérieur, égal ou inférieur à $\varpi_r^{(n)}$. Par suite plusieurs cas pourront se présenter. Ou bien tous les rapports (30) sont plus grands que l'unité ; alors on aura :

$$\varpi_0^{(n)} < \varpi_1^{(n)} < \ldots < \varpi_{n-1}^{(n)} < \varpi_n^{(n)}. \tag{31}$$

Ou bien ils sont tous plus petits que l'unité ; alors

$$\varpi_0^{(n)} > \varpi_1^{(n)} > \ldots > \varpi_n^{(n)}. \tag{32}$$

Ou bien le premier rapport (30) sera plus grand et le dernier plus petit que l'unité ; alors il y aura dans la

suite décroissante (30) deux termes successifs l'un plus grand, l'autre plus petit que l'unité. On aura pour un certain rang R :

$$\frac{\varpi_1^{(n)}}{\varpi_0^{(n)}} > \ldots > \frac{\varpi_R^{(n)}}{\varpi_{R-1}^{(n)}} > 1 > \frac{\varpi_{R+1}^{(n)}}{\varpi_R^{(n)}} > \ldots > \frac{\varpi_n^{(n)}}{\varpi_{n-1}^{(n)}};$$

$$(33)\quad \varpi_0^{(n)} < \varpi_1^{(n)} < \ldots < \varpi_{R-1}^{(n)} < \varpi_R^{(n)} > \varpi_{R+1}^{(n)} > \ldots \varpi_{n-1}^{(n)} > \varpi_n^{(n)}$$

Nous laissons de côté pour le moment le cas exceptionnel où l'un des rapports (30) serait égal exactement à l'unité (il ne peut y en avoir plus d'un dans cette suite décroissante).

Sauf dans ce cas exceptionnel, nous voyons alors que dans les trois cas envisagés il y a un terme $\varpi_R^{(n)}$ plus grand que tous les autres, et que *la suite des $\varpi_r^{(n)}$ va en croissant constamment jusqu'au terme maximum, et ensuite constamment en décroissant, quand r croît.*

Dans le cas exceptionnel où l'un des rapports (30) serait égal à l'unité, ces rapports seront auparavant supérieurs à l'unité et ensuite inférieurs. Donc *s'il y a dans la suite des $\varpi_r^{(n)}$ deux termes égaux supérieurs à tous les autres, cette suite est encore croissante jusqu'au premier de ces deux termes et décroissante depuis le second.*

Rapidité de la décroissance à partir du terme maximum. — Nous avons vu que le rapport de deux termes successifs est

$$\frac{\varpi_{r+1}^{(n)}}{\varpi_r^{(n)}} = \frac{n-r}{r+1}\,\frac{p}{q}.$$

Lorsque r croît de o à $n-1$, ce rapport décroît constamment de $\frac{np}{q}$ à $\frac{p}{nq}$. Si, pour une valeur déterminée de p, n est très grand, la première de ces quantités est très grande, l'autre très petite. On pourra donc choisir deux valeurs r', r'' de r telles que par exemple

$$\frac{n-r'}{r'+1}\frac{p}{q} > 10; \quad \frac{n-r''}{r''+1}\frac{p}{q} < \frac{1}{10}.$$

Il suffit que

$$\frac{n-r'}{r'+1} > \frac{10\,q}{p} \qquad \text{et} \qquad \frac{n-r''}{r''+1} < \frac{q}{10\,p}$$

ou

$$p\,(n-r') > (r'+1)\,10\,q \qquad \text{et} \qquad (n-r'')\,10\,p < q\,(r''+1)$$

ou

$$(33\ bis) \qquad r' < \frac{pn-10\,q}{p+10\,q} \qquad \text{et} \qquad r'' > \frac{10\,np-q}{10\,p+q}.$$

Si n est assez grand ce sera possible et alors : avant le terme de rang $r'+1$ la suite des nombres $\varpi_r^{(n)}$ croîtra plus rapidement qu'une progression géométrique dont la raison est 10 ; après le terme de rang r'', la suite décroîtra plus rapidement qu'une progression géométrique dont la raison est $\frac{1}{10}$. Car on aura :

$$\frac{\varpi_1^{(n)}}{\varpi_0^{(n)}} > \ldots > \frac{\varpi_{r'+1}^{(n)}}{\varpi_{r'}} > 10, \frac{1}{10} > \frac{\varpi_{r''+1}^{(n)}}{\varpi_{r''}^{(n)}} > \ldots > \frac{\varpi_n^{(n)}}{\varpi_{n-1}^{(n)}}.$$

Il en serait de même — à partir d'autres rangs —

si nous avions pris d'autres nombres au lieu de 10 et $\frac{1}{10}$. Donc *quelle que soit la raison d'une progression géométrique, on peut toujours imaginer un nombre* n *d'épreuves assez grand pour que la probabilité* ϖ_r^n *d'une répétition* r *décroisse plus vite que les termes de cette progression* lorsque r croît à partir d'une valeur convenable ou décroît à partir d'une valeur convenable.

Par exemple si on joue à pile ou face 10.000 parties on a $p = q = \frac{1}{2}$, $n = 10.000$, les inégalités (33 *bis*) donnent :

$$r' < \frac{\frac{1}{2} \times 10.000 - 5}{\frac{1}{2} + 5} = \frac{9.990}{11}$$

$$r'' > \frac{10 \times 10.000 \times \frac{1}{2} - \frac{1}{2}}{5 + \frac{1}{2}} = \frac{99.999}{11}.$$

On peut donc prendre par exemple

$$r' = 908\,; \qquad r'' = 9091.$$

Par conséquent dans la suite

$$\varpi_{909}^{(10.000)},\ \varpi_{908}^{(10.000)},\ldots\ \varpi_{1}^{(10.000)},\ \varpi_{0}^{(10.000)}$$

et dans la suite

$$\varpi_{9.091}^{(10.000)},\ \varpi_{9.092}^{(10.000)},\ldots\ \varpi_{9.999}^{(10.000)},\ \varpi_{10.000}^{(10.000)},$$

chaque terme est plus de dix fois plus petit que le précédent.

En résumé, les probabilités $\varpi_r^{(n)}$ non seulement décroissent quand r s'éloigne du rang du terme maximum, mais quand n est assez grand elles décroissent prodigieusement vite quand r est assez éloigné du rang du terme maximum.

Répétition la plus probable. — Ainsi nous venons de vérifier que si n est assez grand il suffit que la répétition r soit assez éloignée du rang R du terme maximum pour que la probabilité $\varpi_r^{(n)}$ de la répétition r devienne très rapidement négligeable.

Il est alors important de connaître R *non pas surtout* parce que, $\varpi_R^{(n)}$ étant la plus grande valeur de $\varpi_r^{(n)}$ quand r varie, *R est la répétition la plus probable*, mais parce que *les valeurs* r *suffisamment éloignées de R deviennent extrêmement improbables* quand n est grand.

Pour préciser, nous désignons par R la répétition la plus probable (ou, si dans la suite des $\varpi_r^{(n)}$ il y a deux termes égaux, la plus grande des deux répétitions les plus probables). Nous aurons alors :

$$\frac{\varpi_R^{(n)}}{\varpi_{R-1}^{(n)}} \geqq 1 > \frac{\varpi_{R+1}^{(n)}}{\varpi_R^{(n)}}$$

d'où, d'après (28) et (29) :

$$\frac{n-R+1}{R}\,\frac{p}{q} \geqq 1 > \frac{n-R}{R+1}\,\frac{p}{q}\,.$$

ou

$$(n-R+1)\,p \geqq Rq \quad \text{et} \quad (R+1)\,q > (n-R)\,p$$

et en remplaçant p par $1 - q$:

$$(n - R + 1)p \geqq R - Rp \quad \text{et} \quad R + 1 - (R + 1)p > (n - R)p;$$

d'où enfin :

$$R \leqq (n + 1)p < R + 1. \tag{34}$$

Le nombre R est entier, $(n + 1)p$ ne l'est pas nécessairement ; d'après (34), R est le plus grand entier qui est contenu dans $(n + 1)p$. Donc :

La répétition la plus probable d'un événement de probabilité p, au cours de n épreuves est égale à la partie entière du produit $(n + 1)p$;

S'il arrive, exceptionnellement, que $(n + 1)p$ soit un nombre entier, ce nombre entier est la plus grande des deux répétitions les plus probables.

Remarquons qu'on a aussi d'après (34)

$$R - 1 < np < R + 1.$$

Par conséquent la partie entière de np est inférieure à $R + 1$ et supérieure ou égale à $R - 1$; elle est donc égale à R ou à $R - 1$.

Nous savons (page 170) que la valeur moyenne de la répétition r dans les groupes de n épreuves est égale à np. On voit donc que la *répétition la plus probable* R *est égale à la répétition moyenne* np *à moins d'une unité près*[1]. En divisant par le nombre fixe n, on voit que : la

1. Il y a lieu d'observer que R peut n'être égal ni à la partie entière, ni à l'entier le plus voisin de np — comme le montre l'exercice I ci-après.

fréquence la plus probable $F = \frac{R}{n}$ est égale à la probabilité p à moins de $\frac{1}{n}$ près.

Il en résulte une manière de préciser la loi du hasard : la probabilité p d'un événement fortuit E est la limite vers laquelle tend la fréquence la plus probable de E dans un groupe de n épreuves lorsque le nombre de ces épreuves croît indéfiniment.

Remarques. — I. — Le procédé qu'on pourrait en déduire pour calculer pratiquement p, serait le suivant. On considérerait un grand nombre N de groupes de n épreuves. On relèverait le nombre N_r de ceux de ces groupes où l'événement de probabilité inconnue p s'est produit r fois. La fréquence $\frac{N_r}{N}$ de ces groupes est une valeur approchée de $\varpi_r^{(n)}$. Si donc N_{R_0} est le plus grand des nombres N_r, $\frac{N_{R_0}}{N}$ sera vraisemblablement voisin de $\varpi_R^{(n)}$, et par suite $\frac{R_0}{n}$ sera voisin de $\frac{R}{n}$, lui-même voisin de p. Ainsi $\frac{R_0}{n}$ sera une valeur empirique approchée de p.

Il y a lieu de remarquer que ce procédé serait de faible valeur. Pour qu'il donne un bon résultat, il faut d'abord que $\frac{R}{n}$ soit voisin de p, et puisque l'on sait seulement que $\frac{R}{n} - p$ est en valeur absolue inférieure à $\frac{1}{n}$, il faudrait que n fut grand. Mais si n est grand il y a un grand nombre de valeurs $\varpi_r^{(n)}$ qui diffèrent

peu de leur maximum $\varpi_R'^{n}$, et, comme en outre $\varpi_r'^{n}$ n'est pas exactement égal à $\frac{N_r}{N}$, il pourra arriver que la plus grande fréquence $\frac{N_{R_0}}{N}$ ait lieu pour une valeur de R_0 notablement différente de R.

II. — Pour de grandes valeurs de r et de n, le calcul direct de l'expression $\varpi_r'^{n}$ au moyen de la formule (26) exigerait la formation du produit d'un très grand nombre de facteurs et deviendrait rapidement impossible. Nous indiquerons plus loin un procédé qui permettrait d'évaluer approximativement $\varpi_r'^{n}$ si n est très grand. Mais il y a lieu de remarquer que si pour une raison quelconque on a pu obtenir la valeur de $\varpi_r'^{n}$ pour une certaine valeur de r, on pourra assez facilement calculer $\varpi_r'^{n}$ pour les valeurs de r voisines de celles-ci. Il suffira de calculer les rapports successifs des diverses valeurs de $\varpi_r'^{n}$ au moyen de la formule (28) qui n'exige que le calcul du quotient de deux nombres entiers donnés.

III. — En établissant la formule (26), nous avons fait usage du théorème des probabilités composées en supposant que le résultat de l'une des n épreuves n'affecte pas la probabilité de l'événement dans les autres épreuves. Il est important d'avoir cette condition présente à l'esprit quand on applique la formule (26). Par exemple, si les n épreuves consistent en n tirages d'une carte d'un paquet, la condition ne sera pas réalisée si on obtient le groupe de n épreuves en tirant d'un coup les n cartes. Et si on les tire successivement, il faudra remettre chaque carte tirée dans le paquet avant

de tirer la suivante. C'est ce qu'on résume en disant que la probabilité doit être constante dans le cours des n épreuves.

EXERCICES

I (F). — Une urne contient 24 boules dont dix-neuf blanches ; ces boules sont identiques à la couleur près Trouver le nombre le plus probable de boules blanches tirées au cours de 22 tirages — chaque boule tirée étant remise dans l'urne avant le tirage suivant. Même problème pour 35 tirages.

II. — Soit ϖ_r la probabilité d'amener r fois pile au cours de dix mille parties du jeu de pile ou face. Montrer que, si r n'est pas trop grand, ϖ_r sera deux fois plus petit que ϖ_{r+1}. Déterminer exactement les valeurs de r pour lesquelles ceci a lieu. Ce sont les entiers inférieurs à un certain nombre r'. Montrer que la probabilité pour que r soit inférieur à r' est inférieure à $\varpi_{r'}$.

III. — Sachant que la probabilité de tirer cent trois fois une boule blanche au cours de mille tirages d'une urne contenant dix boules identiques dont une blanche et neuf noires est égale à 0,0395, calculer la probabilité d'en tirer 99 fois une boule blanche au cours de mille autres tirages (chaque boule tirée étant aussitôt replacée).

CHAPITRE VII

LOIS DES GRANDS NOMBRES

1^re^ Section.

Cas des probabilités constantes. Théorème de Laplace.

Loi des écarts dans le cas des probabilités constantes. Loi-limite de Laplace. — Le théorème de Bernoulli peut être envisagé sous un double aspect. Il ne sert pas seulement à établir la loi du hasard qui, en définitive, est d'avance admise comme vérité d'expérience. Mais il nous apprend quelque chose de plus. Sans déterminer la probabilité P_ε pour que l'écart entre la probabilité fondamentale p et la fréquence $f = \frac{r}{n}$ soit supérieure à un nombre fixe ε, il nous donne une idée de l'ordre de grandeur de P_ε.

Nous n'avons pas reproduit la partie de l'énoncé du théorème original de Bernoulli relative à ce second point, parce qu'après lui on a obtenu à cet égard des résultats plus simples et plus précis que nous allons

indiquer. — Rappelons d'abord que le lemme de Bienaymé nous a suffi pour déterminer une limite supérieure de cette probabilité, par la formule $P_\varepsilon < \frac{\mu'^2}{\varepsilon^2} = \frac{pq}{n\varepsilon^2}$. Des calculs mathématiques dus à Laplace ont permis d'aller plus loin. Sans les développer (ils supposent, en effet, une connaissance assez étendue du calcul intégral) nous nous en tiendrons aux remarques suivantes qui suffisent à établir en quel sens cette loi est vraisemblable, et quelle en est la portée.

Polygones binomiaux. — Considérons un événement E de probabilité p dans une certaine catégorie d'épreuves. Soit un groupe de n épreuves : les probabilités pour que l'événement E s'y répète o fois, 1 fois, 2 fois,..... r,..... n fois ont été appelées (p. 185) respectivement $\varpi_0^{(n)}$, $\varpi_1^{(n)}$, $\varpi_2^{(n)}$,..... $\varpi_r^{(n)}$,..... $\varpi_n^{(n)}$. Nous avons trouvé, après calcul, que, pour un nombre n d'épreuves, les probabilités de cette suite croissent très rapidement jusqu'à un terme maximum et décroissent très vite ensuite. Pour se rendre compte de leur variation, le mieux est de la représenter graphiquement. Portons sur un axe horizontal des longueurs o, $\frac{1}{n}$, $\frac{2}{n}$,..... $\frac{r}{n}$,..... $\frac{n}{n}$ qui mesurent les fréquences possibles, et, sur les verticales passant par ces divisions, les probabilités correspondantes : $\varpi_0^{(n)}$, $\varpi_1^{(n)}$..... On obtiendra ainsi une suite de points : on peut les considérer comme les sommets d'un polygone qu'on appellera le POLYGONE

BINOMIAL *relatif à la probabilité fondamentale* p *et au nombre* n *d'épreuves*[1]. Remarquons qu'il ne dépend que de deux données : le nombre des épreuves, et la probabilité de l'événement. C'est une première propriété très importante : pourvu que ces deux nombres n et p restent les mêmes, il n'y a à tenir compte de la nature ni des épreuves, ni de l'événement : le polygone sera identique.

Mais, en outre, même lorsqu'on fait varier n et p, les polygones obtenus conservent un caractère commun : leurs côtés montent en pente de plus en plus rapide jusqu'au terme maximum (exceptionnellement, s'il y a en cet endroit deux termes égaux, on a alors un côté horizontal), puis descendent suivant une inclinaison de moins en moins forte. Toutefois, quand le nombre n des épreuves n'est pas très grand, les formes de ces polygones sont assez dissemblables, même pour des valeurs p identiques, comme il ressort des figures 7, 8, 9, 10, 11 correspondant aux valeurs successives 4, 9, 25, 100, 1 000 de n, et à la même valeur de p : $p = 0,1$.

Ces différences de forme sont de plus en plus insensibles, et le polygone devient de plus en plus symétrique autour de la verticale du terme maximum, à mesure que n croît.

En revanche, à mesure que n croît, si la forme se stabilise, la grandeur ou la hauteur du polygone dimi-

1. La raison de cette dénomination doit être trouvée à la note[1] de la page 187.

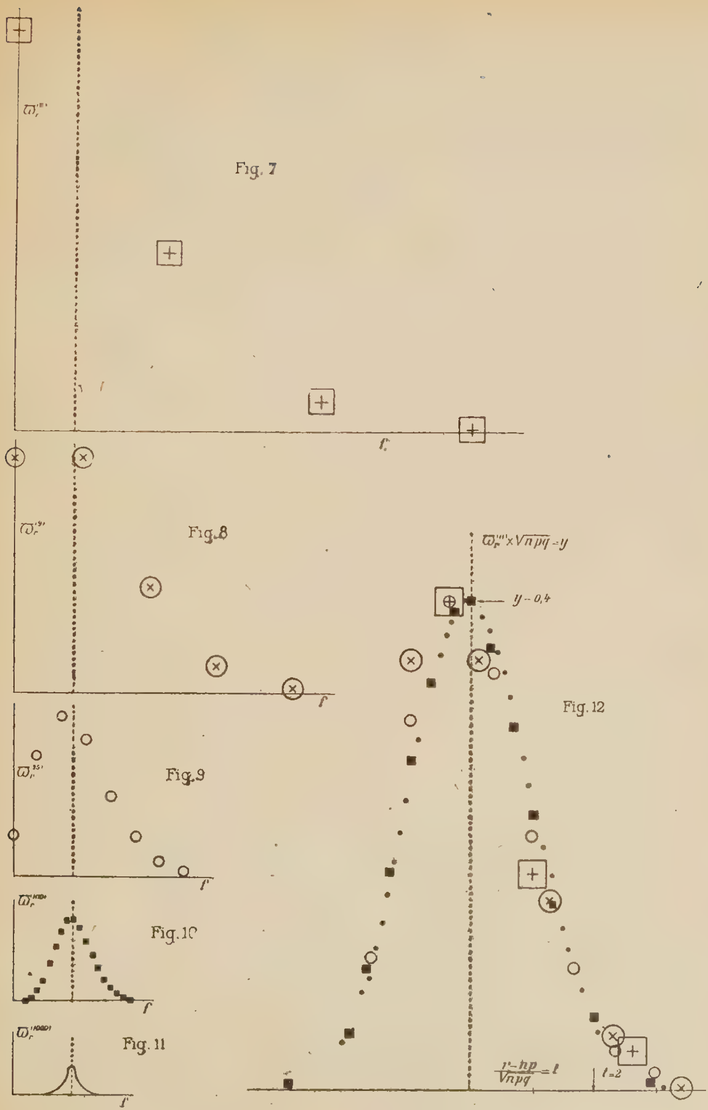


Fig. 7, 8, 9, 10, 11, 12.

nue : le polygone tend à s'écraser sur l'horizontale. Si l'on veut la maintenir constante, il faudra donc changer convenablement les échelles qui servent à mesurer les longueurs horizontales et verticales.

On fixera l'échelle horizontale de la manière suivante : dans le lemme de Bienaymé, nous avons vu que, pour évaluer la probabilité P_ε^n pour que $|f-p| > \varepsilon$, il suffit de connaître non pas ε, mais $\frac{\varepsilon}{\mu'}$ (que nous avons appelé pour cette raison *écart réduit*), en posant $\mu' = \sqrt{\frac{pq}{n}}$. Nous inspirant de ce que la valeur de $\frac{\varepsilon}{\mu'}$ est, en ce sens, plus importante que celle de ε, nous sommes amenés à faire notre changement d'échelle en portant sur l'axe horizontal non pas l'écart absolu, mais l'écart réduit. On peut d'ailleurs calculer aussi bien cet écart réduit : $\frac{f-p}{\mu'}$ en divisant $r - np$ par $\mu = \sqrt{npq}$ (simple transformation algébrique de la formule précédente).

Quant à l'échelle verticale, pour éviter que le polygone ne s'écrase comme précédemment, on la choisira de telle façon que l'aire comprise entre l'axe horizontal et la ligne polygonale ne tende pas vers zéro, et, par exemple, que l'aire limite ait pour mesure l'unité. Lorsque n est très grand, le polygone a un très grand nombre de côtés, et l'aire comprise entre ce polygone et l'axe horizontal est à peu près égale à la somme des aires des rectangles qui ont comme hauteurs les ordon-

nées des sommets du polygone et pour bases les intervalles horizontaux situés entre ces ordonnées. Ces intervalles sont maintenant égaux à $\frac{1}{\mu}$, puisqu'ils sont limités par les points $0, \frac{1}{\mu}, \frac{2}{\mu}, \ldots\ldots \frac{r}{\mu}, \ldots\ldots$ Puisque nous obtiendrons la nouvelle unité de longueur verticale (c'est-à-dire l'échelle verticale nouvelle) en divisant les quantités $\varpi_r^{(n)}$ par une même quantité (encore indéterminée) λ, l'aire cherchée sera approximativement la somme des aires :

$$\frac{1}{\mu} \times \frac{\varpi_r^{(n)}}{\lambda}$$

c'est-à-dire :

$$\frac{1}{\mu\lambda}\left[\varpi_0^{(n)} + \varpi_1^{(n)} + \cdots + \varpi_n^{(n)}\right]$$

Mais la quantité entre crochets est la somme des probabilités pour que, dans un groupe de n épreuves, l'événement E se présente 0 fois, 1 fois,..... n fois ; comme on envisage ainsi tous les cas possibles, cette somme est égale à la probabilité de la certitude, c'est-à-dire à l'unité. L'expression ci-dessus est donc égale à $\frac{1}{\mu\lambda}$. Pour que sa limite soit l'unité, il suffit de prendre $\mu\lambda = 1$, ou $\lambda = \frac{1}{\mu} = \frac{1}{\sqrt{npq}}$. Ainsi on choisira une échelle des longueurs verticales telle que la quantité $\frac{1}{\lambda} = \sqrt{nqp}$ y soit représentée par l'unité de longueur.

Limite des polygones binomiaux dilatés. — On construit ainsi la figure 12 où l'on a représenté les sommets des polygones obtenus par la « dilatation » des polygones binomiaux des figures 7 à 11. Tandis que, sur celles-ci, les polygones, assez dissemblables, révèlent à peine une tendance vers une allure limite, celle-là met en évidence la convergence des polygones dilatés vers une courbe limite de grandeur finie.

Sur la figure 12, qui suffit à illustrer les explications précédentes, sont représentés des polygones binomiaux transformés qui correspondent à une mêms valeur de p : $p = 0,1$. Si on en rapproche les figures obtenues par une transformation analogue de polygones binomiaux qui correspondent à d'autres valeurs de p, on arrive à une conclusion essentiellement distincte des précédentes, et non moins importante : la courbe limite des polygones binomiaux transformés et qui correspondent à un nombre d'épreuves croissant indéfiniment est indépendante de la probabilité fondamentale p.

Loi-limite de Laplace relative à la probabilité d'un écart donné. — Ainsi il existe une courbe bien déterminée (courbe en cloche), — qu'on peut tracer *a priori une fois pour toutes*, — et qui jouit de la propriété suivante : quelle que soit la catégorie d'épreuves, quel que soit l'événement E et sa probabilité p dans cette catégorie, cette courbe représente graphiquement la relation limite entre l'écart de la fréquence avec p et la pro-

babilité de cet écart dans la catégorie d'épreuves envisagée. Supposons qu'on ait tracé les axes horizontaux et verticaux qui déterminent la position de la courbe. On cherche une valeur approchée de la probabilité $\varpi_r^{(n)}$ pour que, dans un groupe de n épreuves d'une catégorie donnée, l'événement E se répète r fois. A cet effet on calcule l'écart quadratique moyen $\mu = \sqrt{npq}$ de la répétition dans ce groupe. On trace une verticale située à une distance de l'axe vertical de symétrie égale au quotient par μ de l'écart $r - np$ entre la répétition envisagée et la valeur moyenne np des répétitions. Le quotient par μ de la longueur interceptée par la courbe en cloche à partir de l'axe horizontal, sur la verticale ainsi tracée, est une valeur approchée de $\varpi_r^{(n)}$. Pour une valeur déterminée de p on peut arriver à une approximation aussi exacte que l'on veut en prenant n assez grand. Telle est la proposition qui a été démontrée mathématiquement par Laplace. Il a établi en même temps la formule — déduite d'une formule de Stirling — qui détermine entièrement la courbe limite.

Vérification expérimentale. — Il est possible d'évaluer $\varpi_r^{(n)}$ par l'observation, ce qui permet de vérifier expérimentalement la loi-limite que nous venons de formuler. On notera, dans un grand nombre N de groupes de n épreuves, le nombre q_r de ceux de ces groupes où l'événement E s'est répété r fois. $\frac{q_r}{N}$ sera la fréquence de la répétition r, et pourra être considérée comme une

valeur approchée de $\varpi_r^{(n)}$. On comparera alors les nombres $\frac{q_0}{N}$, $\frac{q_1}{N}$, ..., $\frac{q_r}{N}$, ... aux nombres correspondants déduits de la courbe en cloche, c'est-à-dire les valeurs observées aux valeurs calculées. S'ils sont voisins, on aura vérifié la loi de Laplace. — Mais il faut remarquer : 1° que $\frac{q_r}{N}$ n'est pas exactement égal à $\varpi_r^{(n)}$, et qu'on ne peut s'attendre à ce qu'ils soient voisins que si N est grand ; 2° que $\varpi_r^{(n)}$ n'est pas exactement égal à la valeur donnée par la courbe en cloche, et qu'on ne peut s'attendre à ce qu'ils soient voisins que si n est grand. — Or les N groupes de n épreuves comportent au total Nn épreuves ; pour arriver à une approximation satisfaisante, il faudra donc observer un nombre énorme d'épreuves. Si par exemple N et n n'atteignent que la valeur modérée d'un millier, Nn correspond déjà à un million d'épreuves. Il sera donc très difficile de vérifier directement la loi de Laplace : ou bien, si l'on cherche une catégorie d'épreuves où la probabilité p ait une valeur bien constante, on recourra aux jeux, et alors on ne parviendra pas à réaliser un nombre d'épreuves suffisant : ou bien, si l'on veut disposer d'un grand nombre d'épreuves, on utilisera des statistiques qui portent sur une nombreuse population : mais alors la probabilité sera sujette à certaines fluctuations.

Exemples numériques. — Nous allons examiner quelques-uns des exemples numériques qui sont cités

dans les traités comme réalisant un bon accord avec la loi limite de Laplace.

1° Czuber a fait avec ses élèves l'expérience suivante.

37 groupes de 100 tirages ont été effectués, d'une urne contenant 6 boules identiques mais numérotées de 1 à 6. (On a donc ici $N = 37$, $n = 100$). A chaque tirage on a noté le numéro de la boule et on a remis celle-ci dans l'urne immédiatement.

On a d'abord considéré comme événement E le tirage d'une boule portant le nombre 1. On a alors noté, pour chaque valeur r de la répétition du tirage de la boule n° 1, le nombre q_r de groupes où on avait constaté cette répétition. On a alors obtenu le tableau suivant :

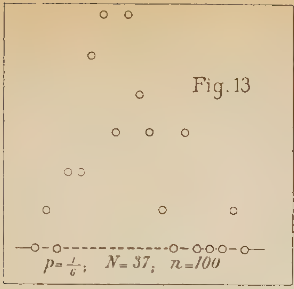
r	q_r	r	q_r
< 10	0	19	3
10	1	20	1
11	0	21	0
12	2	22	3
13	2	23	0
14	5	24	0
15	6	25	0
16	3	26	1
17	6	> 26	0
18	4		
		Total des q_r .	37 groupes

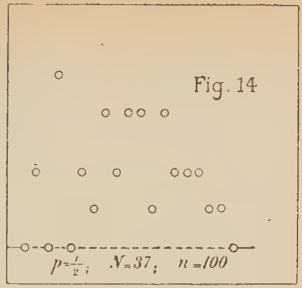
Pour tracer des figures empiriques analogues aux figures théoriques (7-11), il faut marquer sur un axe horizontal les valeurs des fréquences $f = \frac{r}{n}$, soit 0, $\frac{1}{100}$,

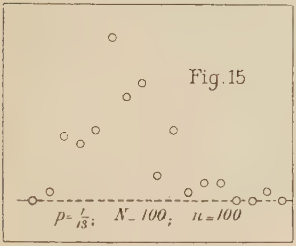
$\frac{2}{100}$, ... $\frac{100}{100}$, et porter sur les verticales les valeurs approchées correspondantes $\frac{q_r}{N}$ de $\varpi_r^{(n)}$. On a alors la figure 13. On voit tout de suite que cette figure ne reproduit qu'en gros la propriété essentielle des polygones binomiaux, savoir que les côtés montent régulièrement et de plus en plus rapidement jusqu'au sommet et descendent ensuite régulièrement et de moins en moins rapidement. Il n'est donc pas nécessaire d'effectuer sur la figure 13 la transformation déjà décrite des polygones pour être assuré que la figure transformée ne sera qu'une très mauvaise approximation de la courbe limite en cloche.

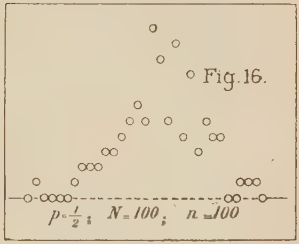
2° Il en est de même lorsque, comme Czuber, on utilise d'une autre façon le relevé des 3.700 tirages, en considérant cette fois comme événement E le tirage d'une boule portant un nombre *pair*. On a alors le tableau suivant :

r	q_r	r	q_r
< 42	0	51	2
42	1	52	4
43	1	53	1
44	2	54	2
45	2	55	0
46	2	56	5
47	4	57	0
48	1	58	2
49	4	> 58	0
50	4		
		TOTAL	37









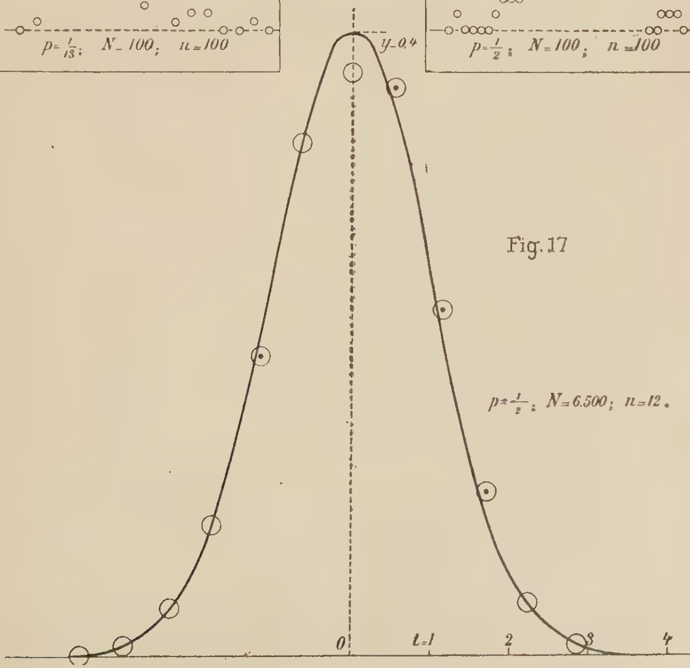


Fig. 13, 14, 15, 16, 17.

La figure 14 correspondante donne lieu aux mêmes remarques que plus haut.

3° Westergaard[1] a effectué 100 groupes de 100 tirages dans une urne contenant un nombre égal de boules blanches et noires, en ayant soin après chaque tirage de remettre la boule tirée, avant le tirage suivant. Il a dressé le tableau suivant où q_r désigne le nombre de groupes où r boules blanches ont été extraites.

$$N = 100, \; n = 100\,;$$

r	q_r	r	q_r
moins de 34	0	50	9
34	1	51	5
35	0	52	10
36	0	53	4
37	0	54	8
38	0	55	3
39	1	56	5
40	2	57	4
41	2	58	4
42	2	59	0
43	3	60	0
44	3	61	1
45	4	62	1
46	5	63	1
47	6	plus de 64	0
48	5		0
49	11		
		Total	100

En procédant comme plus haut on obtient la figure 16 qui elle aussi se rapproche bien grossièrement de la

1. Westergaard, *Die Grundzüge der Theorie der Statistik*, 1890.

courbe en cloche, malgré le nombre imposant (dix mille) des observations faites.

4° Il en est de même de l'expérience faite par Arne Fisher[1] et portant sur 10.000 observations groupées encore en 100 séries de 100 tirages de cartes (replacées immédiatement après chaque tirage dans le paquet de 52 cartes). En considérant comme événement à observer le tirage d'un as, on a obtenu le tableau suivant :

r	q_r	r	q_r
0	0	10	3
1	0	11	9
2	1	12	1
3	8	13	2
4	8	14	2
5	7	15	0
6	9	16	0
7	21	17	1
8	13	> 17	0
9	15		
		TOTAL	100

D'où l'on déduit la figure 15.

Remarque. — Dans les exemples que nous venons de citer, il n'est pas nécessaire, pour faire la figure, de connaître la probabilité fondamentale p. C'est seulement si l'on veut construire le polygone binomial dilaté qu'on a besoin d'évaluer p.

Il est cependant intéressant de noter que ces expériences montrent que la valeur de la probabilité prédite

1. Arne Fisher, *The Mathematical Theory of probabilities*, Macmillan, 1917.

dans les jeux de hasard d'après l'hypothèse des cas également possibles s'accorde assez bien avec l'expérience.

Dans ces expériences, si l'on estime le tirage d'une boule ou d'une autre, celui d'une carte ou d'une autre comme également vraisemblables, les probabilités correspondantes seront évidemment $p = \frac{1}{6}$, $p = \frac{1}{2}$, $p = \frac{1}{2}$, $p = \frac{1}{13}$. D'autre part les tableaux de tirages reproduits plus haut montrent que la fréquence de l'événement dans l'ensemble des tirages était, respectivement : 0,1651 voisin de $p = 0,1666$ ..... ; 0,501 voisin de $p = \frac{1}{2}$ dans la deuxième expérience de Czuber comme dans celle de Westergaard, et enfin 0,0745 au lieu de $p = 0,0769$.

5° Dans les exemples précédents, il suffit de représenter la relation entre $x = \frac{r}{n}$ et $y = \frac{q_r}{N}$, c'est-à-dire de tracer le polygone binomial approché, pour que l'irrégularité de cette relation éclate aux yeux et dispense de passer au polygone transformé. Nous allons indiquer maintenant les résultats d'une expérience due à Darbyshire où le nombre des groupes est assez grand ($N = 6.500$) pour rendre le polygone binomial approché assez régulier. Il faut alors procéder à la transformation de ce polygone pour s'apercevoir que la petitesse de n ($n = 12$) rend encore insuffisante, quoique meilleure, l'approximation par la courbe en cloche. Dans cette expérience, Darbyshire jeta 6.500 fois 12 dés et nota à chaque jet les points obtenus. Il considéra comme

événement favorable l'événement consistant pour un dé à amener l'un des points 4, 5 ou 6 (de sorte qu'en supposant également vraisemblables les arrivées de points quelconques, la probabilité théorique de E serait $\frac{3}{6} = \frac{1}{2}$; en fait la fréquence de E dans les $6.500 \times 12 = 78.000$ jets fut égale à 0,5097).

Il trouva alors

r	q_r	r	q_r
0	1	7	1.351
1	14	8	844
2	103	9	391
3	302	10	117
4	711	11	21
5	1.231	12	3
6	1.411		
		Total	6.500

Il est manifeste que le polygone binomial approché tiré de ce tableau aurait une forme très régulière. Appliquons-lui la transformation déjà décrite, en prenant

$$x = \frac{r - pn}{\mu}\, ; \; y = \left(\frac{q_r}{N}\right) \mu \text{ avec } \mu = \sqrt{npq}$$

où, ici,

$$n = 12,\ N = 6.500,\ p = \frac{1}{2} \cdot$$

On en déduit un tableau de correspondance entre x et y qui permet de construire un polygone dont les sommets sont très voisins d'une courbe en cloche comme

on le voit sur la figure 17 où la ligne en traits pleins est la courbe limite théorique de Laplace. En résumé la loi-limite de Laplace sous la première forme que nous lui avons donnée n'est susceptible d'application numérique que si l'on dispose d'un nombre très considérable d'épreuves.

Probabilité pour que la répétition reste entre deux limites données. — Nous venons de donner à la loi-limite de Laplace une forme qui permettrait d'obtenir une expression approchée de *chacune* des valeurs de $\varpi_r^{(n)}$. Nous allons maintenant résoudre un problème qui permettra de donner à la loi-limite de Laplace une forme comportant une vérification ou prévision numérique beaucoup plus satisfaisante.

On a souvent besoin de connaître non pas la probabilité $\varpi_r^{(n)}$ pour que la répétition ait une valeur donnée, mais la probabilité pour que cette répétition soit comprise entre deux nombres donnés r' et r'', ou, ce qui revient au même, pour que l'écart $l = r - pn$ entre la répétition r et sa valeur moyenne pn soit compris entre deux nombres donnés l' et l''. — On peut considérer la probabilité cherchée comme la somme des probabilités $\varpi_{r_1}^{(n)}$, $\varpi_{r_1+1}^{(n)}$, ... pour que r soit égal à l'un des nombres entiers r_1, ... r_2, compris entre r' et r'', ou entre $l' + np$ et $l'' + np$.

Or nous savons (page 187) calculer la valeur de chacun de ces nombres :

$$\varpi_{r_1}^{(n)} = p^{r_1} q^{n-r_1} C_n^{r_1}, \ldots, \varpi_{r_2}^{(n)} = p^{r_2} q^{n-r_2} C_n^{r_2}$$

On pourra donc calculer leur somme qui est la probabilité cherchée. D'après ces formules, celle-ci est déterminée dès qu'on connaît n, l', l'' et p. On pourra la désigner par $P_{n,p}(l', l'')$.

Mais d'abord, si n est grand, le calcul de C_n^r est très long ; il est vrai qu'alors on peut en obtenir une valeur grossièrement approchée par la courbe en cloche. En outre, et surtout, il faudra peut-être faire le calcul d'un grand nombre de termes $\varpi_r^{(n)}$ s'il y a beaucoup de valeurs de r entre r_1 et r_2. Cet inconvénient ne peut se présenter que si n est grand, puisque r est un entier au plus égal à n. On peut donc essayer d'utiliser *dans ce cas* la courbe en cloche ; même si l'approximation reste assez grossière pour les valeurs de chacun des termes $\varpi_r^{(n)}$ de la somme, comme les erreurs ne sont pas d'un signe déterminé, on peut espérer qu'elles se compenseront au moins partiellement. C'est en effet ce qui se passe, comme il résultera des exemples numériques considérés plus loin. En outre nous allons voir que par ce moyen on peut éviter le calcul séparé même seulement approché d'un grand nombre de termes $\varpi_r^{(n)}$.

Comment opérerons-nous ? Il suffit d'appliquer la règle formulée précédemment (page 207) pour le calcul de $\varpi_r^{(n)}$ à celui de chacun des termes de la somme. Etant donnés la courbe en cloche et les axes qui déterminent sa position, on trace les verticales situées à des distances de l'axe vertical de symétrie de la courbe égales au quotient par $\mu = \sqrt{npq}$ des écarts

$l_1 = r_1 - np$, $l_1 + 1$, ...

Les valeurs approchées de $\varpi_{r_1}^{(n)}, \ldots\ldots \varpi_{r_2}^{(n)}$, sont égales au quotient par μ des longueurs $y_{r_1}, y_{r_1+1}, \ldots\ldots y_{r_2}$ interceptées par la courbe en cloche à partir de l'axe horizontal sur les verticales ainsi tracées. La probabilité cherchée est :

$$P_{n,p}(l', l'') = \varpi_{r_1}^{(n)} + \ldots + \varpi_{r_2}^{(n)} \triangleq \frac{1}{\mu} y_{r_1} + \ldots + \frac{1}{\mu} y_{r_2}.$$

Or l'aire comprise entre la courbe en cloche, l'axe horizontal et les verticales extrêmes est approximativement égale à la somme des rectangles ayant pour hauteurs $y_{r_1}, \ldots\ldots y_{r_2}$, et pour bases les intervalles horizontaux qui les séparent, lesquels sont, d'après ce qui précède, égaux à $\frac{1}{\mu}$. L'approximation est d'autant meilleure que ces intervalles sont plus petits, c'est-à-dire que n est plus grand. L'aire en question ne dépend que des verticales extrêmes, c'est-à-dire de $\frac{l'}{\mu}$ et $\frac{l''}{\mu}$. On peut la désigner par $A\left(\frac{l'}{\mu}, \frac{l''}{\mu}\right)$. On a donc approximativement

$$A\left(\frac{l'}{\mu}, \frac{l''}{\mu}\right) \triangleq \frac{1}{\mu} y_{r_1} + \ldots + \frac{1}{\mu} y_{r_2}$$

d'où

$$P_{n,p}(l', l'') \triangleq A\left(\frac{l'}{\mu}, \frac{l''}{\mu}\right).$$

D'où une seconde forme de l'énoncé de la loi-limite de Laplace, où l'on désignera par t' et t'' les quotients supposés fixes de $\frac{l'}{\mu}$ et de $\frac{l''}{\mu}$:

Loi-limite de Laplace relative à la probabilité d'un écart compris entre deux nombres donnés. — Soient t' et t'' deux nombres fixes (positifs, négatifs ou nuls). La probabilité pour que l'écart $r - np$ entre la répétition r d'un événement E de probabilité constante p au cours de n épreuves, et la moyenne np de r dans ces n épreuves, soit comprise entre t' fois et t'' fois l'écart quadratique moyen $\mu = \sqrt{npq}$ de r dans ces n épreuves tend, lorsque, t', t'' et p restant fixes, n croît indéfiniment, vers l'aire comprise entre la courbe en cloche, la base horizontale de cette courbe, et les deux verticales situées aux distances t', t'' de l'axe vertical de symétrie de la courbe (si t' et t'' sont de signes contraires les verticales seront menées de part et d'autre de l'axe de symétrie.) On remarquera que cette limite est indépendante de p.

Courbe en ogive. — Les applications numériques de l'énoncé précédent présentent des difficultés, car le calcul d'une aire est une opération longue et délicate, même si l'on dispose de l'instrument — le planimètre — destiné à l'effectuer. Il y a avantage à calculer cette aire une fois pour toutes. On peut y parvenir de deux façons différentes. D'abord on représentera graphiquement la variation de cette aire. Désignons par $\psi(\alpha)$ l'aire comprise entre la courbe en cloche, sa base horizontale, et, à droite, une verticale à la distance α de l'axe de symétrie. Il suffit d'examiner les différents cas de figure pour voir que

$$A(t', t'') = \psi(t'') - \psi(t').$$

On remarquera aussi que, pour une valeur déterminée de n, la valeur de cette aire peut être considérée comme une valeur approchée — d'autant plus que n est plus grand — de la probabilité $P_{n,p}(l', l'')$. Or si r_1 et r_2 sont les deux valeurs entières extrêmes comprises entre $l' + np$ et $l'' + np$, cette probabilité ne varie pas si on fait varier l' entre $r_1 - 1$ et r_1, et l'' entre r_2 et $r_2 + 1$. Mais pour pouvoir considérer l'aire cherchée comme somme approchée des aires des rectangles de base $\frac{1}{\mu}$ et de hauteur $y_{r_1}, \ldots\ldots y_{r_2}$, il y aura évidemment avantage à placer r' convenablement entre $r_1 - 1$ et r_1, et r'' entre r_2 et $r_2 + 1$. Sans examiner quelle serait la position exactement la plus avantageuse, il est évident qu'un choix assez satisfaisant consistera à prendre r' à mi-distance entre $r_1 - 1$ et r_1, et r'' à mi-distance entre r_2 et $r_2 + 1$. Ainsi il sera préférable en général de prendre pour r' et r'' des nombres égaux à des entiers augmentés de $\frac{1}{2}$. On fera donc bien de limiter l'emploi de l'égalité approchée

$$A\left(\frac{l'}{\mu}, \frac{l''}{\mu}\right) \simeq P_{n,p}(l', l'')$$

au cas où $r' = l' + np$ et $r'' = l'' + np$ prennent ces valeurs particulières $\left(\text{nombre entier} + \frac{1}{2}\right)$.

Or on peut sur une seconde figure représenter graphiquement la variation de la fonction $\psi(t)$ en portant sur chaque verticale, à la distance t d'un axe vertical

déterminé, une longueur égale à $\psi(t)$ à partir d'une horizontale fixe. L'aire mesurée par $\psi(t)$ va en croissant quand α croît. D'abord très petite, elle reste inférieure à l'aire totale limitée par la courbe en cloche et sa base horizontale, aire totale qu'on a rendue égale à l'unité pour effectuer la transformation des polygones binomiaux. Par conséquent $\psi(t)$ croît constamment de 0 à 1. On obtient alors pour la représentation graphique de la fonction de t une courbe appelée par Galton *courbe en ogive*, et qui est représentée sur la figure 18, page 227. Cette courbe est symétrique par rapport au point O qui correspond à $t = 0$.

On obtient finalement le résultat suivant : I. Quelle que soit la catégorie d'épreuves envisagées, quel que soit l'événement fortuit E, si p est sa probabilité dans cette catégorie, on peut désigner par $P_{n,p}(t'\mu, t''\mu)$ la probabilité pour que l'écart $l = r - np$ entre la répétition r de E dans un groupe de n épreuves et la valeur moyenne np de r dans ce groupe soit comprise entre $t'\mu$ et $t''\mu$, $\mu = \sqrt{npq}$ étant l'écart quadratique moyen de la répétition dans ce groupe. En d'autres termes, cette probabilité ne dépend que de n, p, t', t''. II. *Il existe une courbe dite en ogive qu'on peut dessiner une fois pour toutes, lorsqu'on connaît sa base horizontale et son centre de symétrie O, et qui jouit de la propriété que voici :* lorsque, t', t'' et p restant fixes, n croît indéfiniment, la probabilité $P_{n,p}(t'\mu, t''\mu)$ tend vers la différence des ordonnées $\psi(t'') - \psi(t')$ des points de la courbe en ogive qui sont aux distances t'' et t' de l'axe vertical.

On remarquera que cette limite non seulement ne dépend plus de n, mais est aussi indépendante de p.

On est souvent conduit à rechercher la probabilité pour que l soit, *en valeur absolue*, inférieur à $t\mu$. Cela revient à dire que l doit être compris entre $-t\mu$ et $t\mu$. Or, puisque la courbe en cloche est symétrique, il est évident que la probabilité cherchée est égale au double de la probabilité pour que l soit compris entre o et $t\mu$. Elle a donc pour limite, quand n croît indéfiniment, le double de l'aire $\psi(t) - \psi(o)$, double qu'on peut désigner par $\Psi(t)$. Ainsi la probabilité cherchée tend vers $\Psi(t)$. On peut alors dresser une *table numérique* de la fonction $\Psi(t)$ — (Voir page 290).

Elle donne une expression — d'autant plus approchée que n est plus grand — de la probabilité pour que soit, en valeur absolue, inférieur à t, l'écart *réduit* de la répétition r, c'est-à-dire le quotient, par l'écart quadratique moyen $\mu = \sqrt{npq}$ de r, de l'écart entre r et sa valeur moyenne np.

Il est à remarquer qu'un grand nombre d'auteurs font intervenir dans cet énoncé non la fonction $\Psi(t)$, mais la fonction $\Theta(\lambda)$ qu'on obtient en posant $\Theta(\lambda) = \Psi(\lambda\sqrt{2})$. Autrement dit, au lieu de chercher dans la table de la fonction Ψ la valeur de Ψ qui correspond à $t = \frac{(r - np)}{\mu}$, on cherchera dans celle de Θ la valeur de Θ qui correspond à $\lambda = \frac{(r - np)}{\mu\sqrt{2}}$. On est donc obligé de faire intervenir, à côté de l'écart quadra-

tique moyen μ dont la valeur est en soi utile à connaître, la quantité $\mu\sqrt{2}$ qui n'a aucun intérêt. Au point de vue des calculs numériques qui nous intéresse surtout dans cet ouvrage, l'emploi de la fonction Ψ est préférable. Mais l'expression mathématique de $\Theta(\lambda)$ est légèrement plus simple que celle de $\Psi(t)$, et c'est sans doute la raison qui l'a fait adopter de préférence [1].

Vérification expérimentale. — Par suite de la compensation d'erreurs dont il a été parlé plus haut, la vérification expérimentale de la seconde forme de la loi limite de Laplace est beaucoup plus satisfaisante que celle de la première forme ; en d'autres termes pour la même précision elle n'exige pas un nombre total d'épreuves aussi considérable. On pourra procéder de la façon suivante :

1° *Procédé graphique.* — Étant donnés N groupes de n épreuves, on note encore le nombre q_r de groupes où la répétition d'un même événement fortuit E (de probabilité constante) est égale à r. On en déduit alors le nombre S_X des groupes où la répétition r est inférieure à X ; si R est le plus grand entier inférieur à X, on a en effet :

$$S_X = q_0 + q_1 + \dots + q_R$$

1. Afin de pouvoir la reconnaître et sans qu'il soit utile d'en indiquer ici la signification rappelons l'expression mathématique de $\Theta(\lambda)$:

$$\Theta(\lambda) := \frac{2}{\sqrt{\pi}} \int_0^{\lambda} e^{-x^2}\, dx.$$

Or $\frac{q_0}{N}$, $\frac{q_1}{N}$, ... sont les fréquences des groupes où les répétitions sont o, 1, ... et par suite ces quotients sont les valeurs approchées des probabilités correspondantes $\varpi_0^{(n)}$, $\varpi_1^{(n)}$, ... Donc la quantité $s_X = \frac{S_X}{N}$ est la fréquence des groupes où la répétition est inférieure à X; c'est donc une valeur approchée de la probabilité pour que la répétition r soit inférieure à X. Cette probabilité a pour limite $\psi(t)$, lorsque n croît indéfiniment et que $t = \frac{X - nq}{\sqrt{npq}}$ reste fixe.

Supposons n assez grand pour que la probabilité soit assez voisine de $\psi(t)$; alors si on représente graphiquement la fonction s_X de X, puis si on effectue sur cette courbe la transformation $t = \frac{X - np}{\mu}$, — ce qui revient à déplacer l'origine de la quantité np et à changer l'échelle des longueurs horizontales — on devra obtenir une courbe voisine de la courbe en ogive.

Nous avons vu (p. 220) que pour obtenir un résultat satisfaisant, il vaut mieux ne donner à X que des valeurs égales à un nombre entier augmenté de $\frac{1}{2}$. Par conséquent il vaudra mieux, dans l'énoncé précédent, ne considérer comme voisins de la courbe en ogive que les points de la courbe empirique transformée qui correspondent à ces valeurs de X.

La nécessité d'une limitation de ce genre apparaît immédiatement quand on construit graphiquement une des courbes qui représentent s_X. D'après sa définition

même s_X ne varie pas quand X varie entre deux entiers consécutifs. Si par exemple X est compris entre 13 et 12, S_X reste égal à

$$q_0 + q_1 + \cdot\cdot + q_{12}$$

Donc S_X et par suite aussi $s_X = \frac{S_X}{N}$ reste constant quand X varie entre 13 et 12. Finalement la fonction s_X devra être représentée par une suite de segments horizontaux s'élevant graduellement de la hauteur 0 à la hauteur 1. La transformation $t = \frac{X - np}{\mu}$ altèrera la longueur de ces segments et leur position, mais la courbe transformée sera encore constituée de segments horizontaux de longueurs égales à $\frac{1}{\mu}$ et dont les hauteurs s'élèvent graduellement de 0 à 1. Il est alors certain par avance que si $\frac{1}{\mu}$ n'est pas très petit (c'est-à-dire si n n'est pas très grand) cette suite discontinue de segments parallèles ne pourra être partout voisine de la courbe en ogive. Il n'en est plus de même si l'on envisage le voisinage avec la courbe en ogive des milieux de ces segments. C'est ce que manifeste par exemple la figure 18, obtenue en appliquant les considérations précédentes aux exemples numériques traités précédemment. Ils montrent que, dans chaque cas, la courbe en ogive n'est que grossièrement voisine de la courbe transformée de la courbe représentant s_X dans toute son étendue, mais que, limitée aux milieux des

segments, l'approximation est beaucoup plus satisfaisante. En particulier la suite des milieux des segments reproduit l'allure générale de la courbe en ogive d'une façon beaucoup plus satisfaisante que les figures 7-11 ne reproduisaient la courbe en cloche.

En résumé si l'on veut s'assurer si un événement fortuit E obéit aux lois du hasard, et si l'on trace les courbes empiriques qui doivent représenter approximativement l'une la courbe en cloche, l'autre la courbe en ogive on doit s'attendre à voir reproduite l'allure générale et même la grandeur de la courbe en ogive beaucoup mieux que celle de la courbe en cloche.

En d'autres termes, en utilisant la loi-limite de Laplace, on prévoit quelle proportion parmi un certain nombre de groupes de n épreuves présentent une répétition de l'événement E comprise entre deux nombres assez distants plus exactement que la proportion de ceux où cette répétition a une valeur donnée.

2° *Procédé numérique.* — En général, quand on étudie la probabilité d'un écart compris entre deux nombres donnés, on s'en tient au cas où ces deux nombres sont égaux et de signes contraires. En d'autres termes, on se propose d'évaluer la proportion $\frac{Q}{N}$ de ceux des N groupes de n épreuves où l'écart entre la répétition de l'événement dans ces épreuves avec la valeur moyenne de cette répétition est, *en valeur absolue*, inférieure à un nombre donné. Nous avons vu que la probabilité pour que $| r - np | < t\mu$, où $\mu = \sqrt{npq}$,

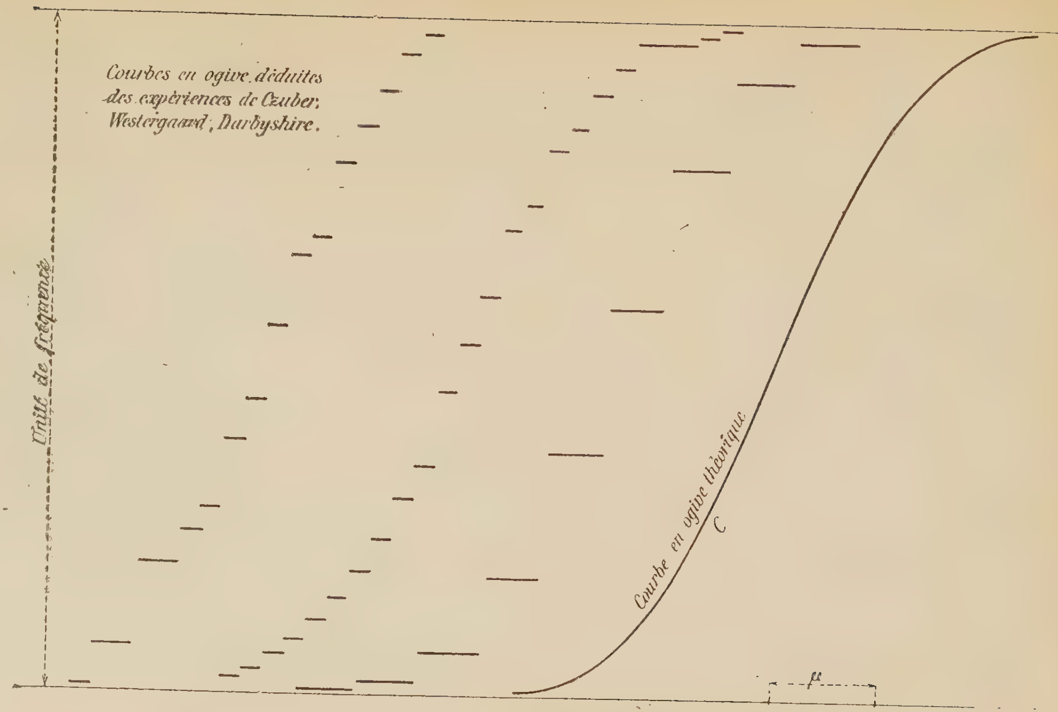

Courbes en ogive déduites
des expériences de Czuber;
Westergaard; Darbyshire.
Unité de fréquence
Courbe en ogive théorique
C
μ

Fig. 18.

a pour valeur approchée $\Psi(t)$ quand n est suffisamment grand ; d'autre part, si N est grand, la fréquence $\frac{Q}{N}$ est en général voisine de cette probabilité.

Si donc N et n sont assez grands, on aura approximativement :

$$\frac{Q}{N} \doteq \Psi\left(\frac{l}{\mu}\right)$$

de sorte que la valeur approximative cherchée pour Q sera :

$$N\Psi\left(\frac{l}{\mu}\right)$$

où la valeur de Ψ est donnée par la table (p. 290) en face de $t = \frac{l}{\mu} = \frac{r - np}{\sqrt{npq}}$.

Les trois causes d'erreur. — Lorsqu'on cherche à évaluer Q par ce procédé, on se trouve exposé à trois causes d'erreur. Il importe de les connaître, pour ne pas être surpris si la valeur effective de Q paraît très différente de celle qu'on prévoyait, et aussi afin, quand cela est possible, de réduire ces divergences par des méthodes appropriées.

1° Comme $\frac{Q}{N}$ est une fréquence, celle-ci peut être très et même exceptionnellement différente de la probabilité correspondante. Si l'on prend N assez grand on diminuera beaucoup les chances pour qu'une telle difrence intervienne.

2° La probabilité envisagée pourra s'écarter de $\Psi(t)$ qui n'est que la limite de cette probabilité quand n

croît indéfiniment. On réduira donc cette cause d'erreur en prenant n aussi grand qu'il est pratiquement possible.

3° Lorsqu'on donne au nombre r toutes les valeurs entières possibles entre o et n, on trouve pour $|r - np|$ un nombre fini de valeurs. Supposons ces valeurs absolues rangées par ordre de grandeur croissante : l_1, l_2, ... Si on fait varier l entre deux de ces valeurs successives l_k et l_{k+1}, le nombre Q ne varie pas. Or, la règle précédente donne, comme valeur de Q, un nombre $N\Psi\left(\frac{l}{\mu}\right)$ qui varie avec l. Mais il est impossible que Q soit égal à la fois à toutes ces valeurs de l.

Comme celles-ci sont toutes voisines de Q si n est extrêmement grand, il est naturel d'admettre que, si n est modérément grand, Q sera compris entre la plus grande et la plus petite de ces valeurs qui sont $N\Psi\left(\frac{l_k}{\mu}\right)$ et $N\Psi\left(\frac{l_{k+1}}{\mu}\right)$. Et par conséquent il sera plus prudent, pour évaluer Q, de prendre pour l la moyenne arithmétique de l_k et l_{k+1}. Dans le cas où np est un nombre entier, ce ne serait qu'une conséquence de la remarque analogue faite à la page 224 et qui repose sur l'évaluation des aires. Si np n'est pas entier, la règle que nous venons d'indiquer est distincte de la précédente, et on en peut donner une justification mathématique plus précise que celle qui nous a suffi [1].

1. Voir Castelnuovo, *Calcolo delle Probabilita*, page 91. L'intro-

A titre d'application, reprenons l'exemple de Czuber de la page 210.

On a ici : $p = \frac{1}{2}$, $n = 100$, $N = 37$; d'où :

$$\mu = \sqrt{npq} = \sqrt{100 \times \frac{1}{2} \times \frac{1}{2}} = 5.$$

Alors si Q est le nombre des groupes de 100 épreuves où $|r - np| < l$, on a approximativement :

$$Q \simeq N\Psi\left(\frac{l}{\mu}\right) = 37\,\Psi\left(\frac{l}{5}\right).$$

Les écarts possibles de l'entier r avec sa moyenne $np = 50$ sont 0, 1, 2, ... de sorte que, conformément à la règle indiquée plus haut nous prendrons :

$$l = \frac{1}{2},\ \frac{3}{2},\ \frac{5}{2},\ \ldots.$$

et par suite nous calculerons $\Psi(t)$ pour $t = \frac{l}{5} = 0{,}1$; 0,3 ; 0,5 ; ..... Nous indiquerons alors sur une ligne les valeurs de Q correspondantes, et sur une troisième ligne les valeurs approchées de Q au moyen de l'expression $37\,\Psi\left(\frac{l}{5}\right)$ sous le nom de Q calculé. Nous nous contenterons de faire le calcul pour

$$l = \frac{5}{2},\ \frac{9}{2},\ \frac{13}{2},\ \frac{17}{2}$$

duction de $\frac{l_k + l_{k+1}}{2}$ revient à ajouter à $\Psi\left(\frac{l_k}{\mu}\right)$ un terme complémentaire (considéré par Laplace) qui tend vers zéro quand n croît, mais qui donne une meilleure approximation.

Ecart l	$\frac{5}{2}$	$\frac{9}{2}$	$\frac{13}{2}$	$\frac{17}{2}$
Q	15	24	33	37
Q calculé	14,2	23,4	29,8	33,7

Étant donné le petit nombre (37) des groupes, on peut considérer l'accord comme satisfaisant.

Dans la suite nous pourrons essayer de chiffrer les trois erreurs que nous avons introduites. En ce qui concerne la troisième, on voit tout de suite qu'au lieu d'écrire que la valeur de Q pour l entre 0 et 0,04 est 19,1, on peut dire qu'elle est comprise entre 21,3 et 23,4 ; la valeur observée est 24. — En ce qui concerne la seconde erreur on a remplacé la probabilité P_n par la fréquence $\frac{Q}{N}$. Quel est l'écart quadratique moyen de P_n dans l'ensemble des N groupes. C'est :

$$\mu' = \sqrt{\frac{P_n\,(1 - P_n)}{N}}\,.$$

On peut alors chercher quelle est l'erreur commise sur P_n telle que la probabilité d'une erreur plus grande ne serait que de $\frac{1}{100}$. Ce serait une valeur α telle que $\frac{\alpha}{\mu'} < t$, où t donne pour $\Psi\,(t)$ une valeur 0,99, soit, d'après la table de Ψ, $t = 2{,}6$; donc ;

$$\frac{\alpha}{\mu'} < 2{,}6 ;\ \alpha < 2{,}6\sqrt{\frac{P_n\,(1 - P_n)}{N}}$$

ou, si l'on suppose par exemple $l = \frac{3}{2}$, en remplaçant approximativement P_n par la valeur $\Psi = 0,236$:

$$\alpha < 2,6\sqrt{\frac{0,236 \times 0,764}{37}} < 2,6\sqrt{0,00487} < 2,6 \times 0,07 = 0,182.$$

Ainsi donc il y a 99 à parier contre un que la seconde erreur commise en prenant, pour $\frac{Q}{N}$, P_n, est inférieure à 0,182. L'erreur commise de ce fait sur $Q = 37\ P$ serait donc $< 0,182 \times 37 = 6,73$.

Si l'on est moins exigeant sur la certitude du résultat, on peut réduire ainsi l'erreur : cherchons l'erreur dite probable commise sur P_n.

On a $\Psi = \frac{1}{2}$ pour $\frac{\alpha}{\mu'} = 0,674$ environ ; alors α sera égal à $0,674\sqrt{0,00487}$ = environ 0,047. Donc il y a un contre un à parier que l'erreur sur P_n sera inférieure à 0,05 ; c'est ce qu'on exprime en disant que 0,05 est l'erreur probable sur P_n. L'erreur probable sur Q sera 1,7.

Représentation analytique approchée de la loi des écarts. — Nous savons que, pour un très grand nombre d'épreuves, la probabilité d'un écart absolu supérieur à t fois l'écart quadratique moyen μ est sensiblement égal à $1 - \Psi(t)$.

Pour la commodité du calcul nous allons voir qu'il peut être pratiquement plus commode d'évaluer la probabilité π_k d'un écart absolu $| l |$ supérieur à k fois

l'*écart décimal* en appelant ainsi, d'après M. Borel, l'écart l_0 tel que la probabilité d'un écart absolu supérieur soit égale à $\frac{1}{10}$. On aura :

$$1 - \Psi\left(\frac{l_0}{\mu}\right) = \frac{1}{10}$$

et

$$\pi_k = 1 - \Psi(t) \quad \text{si} \quad kl_0 = t\mu$$

d'où :

$$\pi_k = 1 - \Psi\left(k\,\frac{l_0}{\mu}\right).$$

On a par définition de π_k, $\pi_1 = \frac{1}{10}$. Cherchons de même les valeurs de k pour lesquelles π_k prend les valeurs $1, \frac{1}{10}, \frac{1}{100}, \frac{1}{1000}, \ldots\ldots$ En posant $\pi_{k_r} = \frac{1}{10^r}$, on aura :

$$\Psi\left(k_r\,\frac{l_0}{\mu}\right) = 1 - \frac{1}{10^r},$$

et en posant $t_r = k_r\,\frac{l_0}{\mu}$

$\Psi(t_0) = 1$; $\Psi(t_1) = 0{,}9$; $\Psi(t_2) = 0{,}99$; $\Psi(t_3) = 0{,}999$; ..... ; $\Psi(t_8) = 0{,}99999999$.

En se reportant à la table de la fonction Ψ, on trouve, à un centième près,

$t_0 = 0$; $t_1 = 1{,}64$; $t_2 = 2{,}58$; $t_3 = 3{,}29$; $t_4 = 3{,}89$, $t_5 = 4{,}42$; $t_6 = 4{,}89$; $t_7 = 5{,}33$; $t_8 = 5{,}73$.

Et comme :

$$\frac{t_r}{k_r} = \frac{t_0}{\mu} = \frac{t_1}{k_1}$$

et $k_1 = 1$, d'où $k_r = \frac{t_r}{t_1}$, on trouve :

$k_0 = 0$, $k_1 = 1$, $k_2 = 1{,}57$; $k_3 = 2{,}01$; $k_4 = 2{,}38$, $k_5 = 2{,}70$; $k_6 = 2{,}99$; $k_7 = 3{,}25$; $k_8 = 3{,}5$.

Si on porte horizontalement les nombres k_r et sur les verticales correspondantes les nombres r, on voit que la suite de points obtenus dessine à peu près une parabole.

Ceci suggère que l'on pourrait représenter approximativement r par une fonction du second degré de k_r.

Puisque $k_3 = 2{,}01$ et $k_6 = 2{,}99$ sont approximativement les entiers 2 et 3, on voit que cette fonction du second degré sera égale à 0, 1, 3, 6 pour $k = 0, 1, 2, 3$. Ce sera donc $r = \frac{k(k+1)}{2}$. Vérifions en calculant $\frac{k(k+1)}{2}$ pour les autres valeurs de la suite précédente ; on a le tableau :

k	0	1	1,57	2,01	2,38	2,70	2,99	3,25	3,5
r	0	1	2	3	4	5	6	7	8
$\frac{k(k+1)}{2}$	0	1	2,01	3,02	4,02	4,99	5,96	6,91	7,87

Ainsi on a approximativement :

$$r \simeq \frac{k_r(k_r+1)}{2}$$

et d'autre part $\pi_{k_r} = \frac{1}{10^r}$, d'où :

$$\pi_{k_r} \backsimeq \frac{1}{10^{\frac{k_r(k_r+1)}{2}}}$$

et en général[1]

$$\pi_k = \frac{1}{10^{\frac{k(k+1)}{2}}}.$$

Ainsi nous arrivons à une expression approchée très simple de la loi des écarts :

Dans un grand nombre d'épreuves la probabilité π_k *pour que l'écart absolu* $|\,l\,|$ *surpasse* k *fois l'écart décimal est approximativement égal à l'inverse d'une puissance de* 10 *dont l'exposant est* $\frac{k(k+1)}{2}$.

A vrai dire nos raisonnements ne s'appliquent que pour les valeurs de k considérées plus haut, pour lesquelles l'exposant de 10 est entier. Pour aller au delà il serait d'ailleurs nécessaire de supposer la connaissance de la théorie des exposants fractionnaires ou irrationnels, ce que nous désirons éviter. Disons toutefois que l'expression que nous venons d'obtenir fournit aussi une bonne approximation quand l'exposant n'est pas entier.

1. Voir Fréchet, *Revue générale des Sciences*, page 211, 1923. L'idée d'établir une formule de ce genre nous a été inspirée par une formule analogue donnée par M. Borel, formule moins précise que la nôtre mais plus simple, où $\frac{k(k+1)}{2}$ est remplacé par k.

Remarquons que l'on a $t_r = t_1\, k_r = 1,64\, k_r$, d sorte qu'on peut encore écrire :

$$1 - \Psi(1,64\, k) \simeq \frac{1}{10^{\frac{k(k+1)}{2}}}.$$

Comme vérification on peut dresser le tableau suivant :

t	0	0,16	0,64	1	1,6448	2,59	4
$1 - \Psi(t)$	1	0,87	0,52	0,32	0,1	0,0096	0,000063
$\frac{1}{10^{\frac{k(k+1)}{2}}}$	1	0,88	0,54	0,32	0,1	0,0094	0,000067

L'expression de la loi des écarts que nous venons de formuler peut rendre des services quand il s'agit d'apprécier rapidement la probabilité d'un écart sans avoir besoin de recourir à une table qu'on n'a pas toujours sous la main. Elle a aussi l'avantage de mettre en évidence la rapidité de la décroissance de la probabilité d'un écart croissant, $\frac{k(k+1)}{2}$ donnant généralement le rang de la première décimale de cette probabilité.

Loi des petites probabilités. — Nous avons déjà calculé (p. 187) la valeur de la probabilité $\varpi_r^{(n)}$ pour que dans un groupe de n épreuves, la répétition d'un événement de probabilité p soit égale à r.

Lorsque n est très grand, le calcul très long de $\varpi_r^{(n)}$

peut être abrégé par l'emploi de la courbe en cloche. Mais nous avons remarqué que le polygone binomial se rapprochait d'autant plus vite de la courbe en cloche, quand n croît, que p est moins éloigné de $\frac{1}{2}$. La loi-limite de Laplace ne donnera pas au contraire une bonne approximation quand p est éloigné de $\frac{1}{2}$, voisin de 0 ou 1, tant que n n'est pas extrêmement grand. Il est donc utile d'indiquer dans ce cas une expression limite de la loi des écarts ; on formulera ainsi ce qu'on peut appeler la loi-limite des petites probabilités — loi qui a été découverte par Poisson.

Dans ce but rappelons que :

$$\frac{\varpi_r^{(n)}}{\varpi_{r-1}^{(n)}} = \frac{n-r+1}{r}\,\frac{p}{q} = \frac{1-\dfrac{r-1}{n}}{r}\,\frac{np}{q}.$$

Appelons R la quantité np (qui est la valeur moyenne de r) et supposons que, p étant très petit, on se contente d'évaluer celles des valeurs $\varpi_r^{(n)}$ pour lesquelles r est notablement inférieur à sa valeur probable et par suite à R. Alors $\frac{r-1}{n}$ est petit, $q = 1 - p$ est voisin de 1, et l'on a approximativement :

$$\frac{\varpi_r^{(n)}}{\varpi_{r-1}^{(n)}} = \frac{R}{r}$$

Par conséquent :

$$\frac{\varpi_1^{(n)}}{\varpi_0^{(n)}} \triangleq R,\quad \frac{\varpi_2^{(n)}}{\varpi_1^{(n)}} \triangleq \frac{R}{2},\ \ldots\ \frac{\varpi_r^{(n)}}{\varpi_{r-1}^{(n)}} \triangleq \frac{R}{r}.$$

D'où (en multipliant ces égalités) :

$$\frac{\varpi_r^{(n)}}{\varpi_0^{(n)}} \triangleq \frac{R^r}{r\,!}$$

en posant $r\,! = 1 \times 2 \times \ldots\ldots r$; et par suite :

$$(35\ ter) \qquad \frac{\varpi_0^{(n)}}{1} \triangleq \frac{\varpi_1^{(n)}}{R} \triangleq \frac{\varpi_2^{(n)}}{\frac{R^2}{2}} \cdots \triangleq \frac{\varpi_r^{(n)}}{\frac{R^r}{r\,!}} \cdot$$

Donc *si* p *est petit les probabilités* $\varpi_0^{(n)}$, $\varpi_1^{(n)}$, ... $\varpi_r^{(n)}$, ..., *pour que la répétition* r *d'un événement de probabilité* p, *dans un groupe de* n *épreuves, soit égale respectivement à* 0, 1, 2 ..... r *sont approximativement proportionnelles aux nombres* :

$$(36) \qquad 1,\ R,\ \frac{R^2}{2},\ \ldots\ \frac{R^r}{r\,!}$$

où R est la valeur moyenne np *de* r.

Il ne reste plus qu'à trouver le coefficient de proportionnalité, c'est-à-dire la valeur approximativement commune des rapports (35 *ter*). Poisson l'a déterminé en évaluant une valeur approchée de $\varpi_0^{(n)}$ — qui est égal exactement comme on sait (page 187) à q^n. Il remplace q^n par l'expression $(1 - p)^{\frac{R}{p}}$ et prend comme expression approchée de celle-ci, pour les petites valeurs de p,

la limite de cette expression quand p tend vers zéro, R restant fixe[1].

Sans avoir besoin de connaître cette limite, nous remarquons que sa valeur ne dépend plus que de R. Si donc nous tirons de (35 *ter*)

$$\varpi_1^{(n)} \doteq R\,\varpi_0^{(n)}, \qquad \varpi_2^{(n)} \doteq \frac{R^2}{2}\,\varpi_0^{(n)}, \ldots$$

et si nous remplaçons $\varpi_0^{(n)}$ par cette valeur approchée qui ne dépend plus que de R, nous parvenons au résultat suivant.

Quelle que soit la probabilité p *d'un événement fortuit E, la loi de Poisson permet d'assigner à la probabilité* $\varpi_r^{(n)}$ *que la répétition de E dans* n *épreuves soit égale à* r, *une valeur approchée qui ne dépend que de* r *et de la valeur moyenne* $R = np$ *de* r, *sans qu'il soit nécessaire de connaître ni* p, *ni* n *pourvu que* p *soit très petit et que l'on se borne aux valeurs de* r *notablement plus petites que* R.

Il est donc possible d'établir des tables numériques à double entrée donnant pour chaque valeur de r les valeurs approchées de $\varpi_r^{(n)}$ correspondant à diverses valeurs de R ou inversement.

Mais nous allons indiquer un autre moyen de cal-

1. Tout lecteur familier avec la théorie des exponentielles reconnaîtra que :

$$\lim_{p \to 0} (1-p)^{\frac{R}{p}} = \lim_{p \to 0} \left[\frac{1}{(1-p)^{-\frac{1}{p}}}\right]^R = \frac{1}{e^R} = e^{-R}.$$

Mais nous n'utiliserons pas cette égalité.

culer le coefficient de proportionnalité, moyen qui nous dispensera de l'usage de ces tables sans donner lieu à des calculs beaucoup plus longs, et qui nous permettra d'établir la loi que nous venons d'énoncer sans recourir à la théorie des exponentielles. Examinons, pour plus de simplicité, comment on s'en sert dans la vérification expérimentale.

Vérification expérimentale. — Nous prendrons un grand nombre N de groupes de n épreuves et nous noterons le nombre P_r de ces groupes où la répétition d'un certain événement fortuit E est r. Il s'agit de calculer, au moyen de la loi des petites probabilités, des valeurs approchées P'_r de P_r, et *cela sans connaître* n.

On connaît le nombre total S d'épreuves où E s'est produit; il s'agit d'en prévoir la répartition entre les N groupes. Le rapport $\frac{S}{N}$ donne une valeur approchée ρ de la valeur moyenne R de la répétition dans un groupe de n épreuves. Les fréquences observées $\frac{P_0}{N}$, $\frac{P_1}{N}$, ... des valeurs $0, 1, \ldots$ de la répétition sont des valeurs approchées de $\varpi_0^{(n)}$, $\varpi_1^{(n)}$, ...

Ces dernières doivent être, d'après 35 *ter*, approximativement proportionnelles à 1; $\frac{R}{1}$; $\frac{R^2}{2}$; $\frac{R^3}{6}$; ....., lesquelles sont approximativement égales à 1; $\frac{\rho}{1}$; $\frac{\rho^2}{2}$; $\frac{\rho^3}{6}$; .....

Par conséquent $P_0, P_1, P_2, \ldots\ldots$ doivent être approximativement proportionnels à $1, \frac{\rho}{1}, \frac{\rho^2}{2}, \ldots\ldots$ On a donc pour les valeurs approchées $P'_0\ P'_1\ P'_2, \ldots\ldots$ qu'il s'agit de calculer.

$$\frac{P'_0}{1} = \frac{P'_1}{\rho} = \frac{P'_2}{\frac{\rho^2}{2}} = \frac{P'_3}{\frac{\rho^3}{6}} = \ldots = \frac{P'_r}{\frac{\rho^r}{r!}} = \ldots$$

Il ne reste plus qu'à calculer le coefficient de proportionnalité k. Or, la suite des dénominateurs, $\frac{\rho^r}{r!}$ décroît très rapidement dès que r est notablement plus grand que ρ puisque le terme $\frac{\rho^r}{r!}$ est égal au précédent $\frac{\rho^{r-1}}{(r-1)!}$ multiplié par le rapport $\frac{\rho}{r}$ qui sera petit.

Par suite dès que r sera assez grand les dénominateur deviendront négligeables, et par suite aussi les numérateurs P'_r. Si, par exemple, on constate qu'à partir du terme de rang r' les dénominateurs sont négligeables, on pourra écrire

$$P'_0 + P'_1 + \ldots + P'_{r'} \simeq P'_0 + P'_1 + \ldots + P'_{r'} + \ldots \simeq P_0 + P_1 + \ldots = N$$

et

$$P'_0 = k,\ P'_1 = k\rho, \ldots, P'_{r'} = k\,\frac{\rho^{r'}}{r'!},$$

d'où :

$$N \simeq k\left(1 + \rho + \ldots + \frac{\rho^{r'}}{r'!}\right).$$

D'où on tirera la valeur de k et par suite les valeurs P'_0, P'_1, ... [1], sans connaître ni n, ni p séparément.

Exemple. — Dans une statistique dressée par M. Huber, Directeur de la Statistique générale de la France [2], on trouve un tableau donnant par département le nombre de décès par jour dans un groupe de mille enfants d'un an placés en nourrice. Le relevé a été fait dans les 37 départements où le nombre des enfants observés dépasse mille.

Comme les nombres journaliers obtenus sont inférieurs à l'unité, nous avons admis pour simplifier que l'on peut en déduire les nombres de décès par mois de 30 jours en multipliant les nombres indiqués par 30 et arrondissant en nombres entiers. Un simple comptage permet alors de relever le nombre P_r de ceux des 37 départements où le nombre de décès par mois et par mille enfants est égale à r.

Nous voulons comparer aux nombres P_r observés les nombres P'_r que fournirait la loi des petites probabilités si l'on connaissait seulement le relevé total pour l'ensemble des 37 départements et si l'on voulait en déduire la répartition entre ces départements.

La valeur moyenne ρ de r déduite de ce relevé total étant 2,45, il est d'abord facile de calculer les valeurs

1. La théorie de l'exponentielle permettrait d'éviter au moyen d'une table appropriée le calcul de la somme entre parenthèses égale approximativement, d'après, cette théorie, à la quantité désignée par e^{ρ}, d'où l'on déduirait $k = Ne^{-\rho}$, ce qui ramène à la formule de Poisson.

2. *Bull. de l'Institut intern. de Statistique.*, t. XVIII, 1909.

des nombres 1, ρ, $\frac{\rho^2}{2}$, ... $\frac{\rho^r}{r!}$, ... On a le tableau

r	0	1	2	3	4	5	6	7	8	TOTAL
$\frac{\rho^r}{r!}$	1	2,45	3,00	2,45	1,50	0,73	0,30	0,10	0,03	11,56

où nous avons laissé de côté les valeurs de r supérieures à 8 qui fournirait des valeurs $\frac{\rho^r}{r!}$ inférieures à un centième.

Ceci étant les valeurs P'_r à calculer doivent fournir un total voisin du nombre total $N = 37$ des départements observés, et elles doivent être proportionnelles aux nombres de la deuxième ligne du tableau ci-dessus, dont le total est 11,56.

Par suite, on obtiendra ces valeurs P'_r en multipliant les nombres de cette seconde ligne par un coefficient de proportionnalité k égal à $\frac{37}{11{,}56} = 3{,}20$, et en arrondissant à une demi-unité près. On obtient alors les nombres $P'_r = k\frac{\rho^r}{r!}$, que nous avons placés dans le tableau ci-dessous sous le nom de P'_r calculé au-dessous des nombres P_r réellement observés, déduits de la statistique de M. Huber :

r	0	1	2	3	4	5	6	7	8	TOTAL
P_r observé	0	8	14	8	5	1	1	0	0	37
P'_r calculé	3	8	10	8	5	2	1	0	0	37

On peut considérer l'accord entre les valeurs observées et calculées comme satisfaisant, étant donné surtout le petit nombre $N = 37$ des groupes observés.

Remarque. — Ce qui rend cette loi particulièrement précieuse, c'est que les valeurs de n et de p n'y interviennent pas directement. Or, dans nombre de statistiques, on connaît le nombre de répétitions d'un événement assez rare sans savoir d'une façon très précise, sur quel groupe il porte. Par exemple le nombre d'accidents mortels survenus dans certaines professions, où le nombre des ouvriers varie dans l'année tandis que le nombre annuel de journées varie dans de moindres proportions mais n'est que très vaguement connu.

EXERCICE

I (F). — Durant 26 années, on a observé la fréquence f des garçons parmi les mort-nés. Comme il y a eu en moyenne chaque année 10.319 garçons parmi 17.956 mort-nés, on peut admettre pour valeur de la probabilité correspondante $p = \frac{10.319}{17.956} = 0{,}5747$. On a alors calculé les écarts annuels $f - p$ qui deviennent en les multipliant par 10.000 pour éviter les décimales

$+ 12$, $- 11$, $+ 36$, $+ 58$; $- 27$; $+ 34$, $+ 78$, $- 24$, $- 39$; $- 28$; $+ 4$ $- 31$; $+ 56$, $- 65$; $+ 11$, $+ 9$; $- 17$; $- 13$; $- 72$; $+ 22$, $- 41$; $+ 32$; $- 36$; $+ 9$; $+ 10$; $+ 33$;

Tracer la courbe empirique correspondant à cette loi des écarts ; celle où à chaque valeur de x correspond une

ordonnée S_x égale au nombre d'années où l'écart a été inférieur à x.

Tracer la courbe en ogive empirique correspondante obtenue en remplaçant l'écart $f-p$, par l'écart réduit $\frac{f-p}{\mu}$ où $\mu = \sqrt{\frac{p(1-p)}{n}}$. On admettra que le nombre des mort-nés a varié assez peu d'une année à l'autre pour qu'on puisse supposer n constant et égal à sa valeur moyenne 17.956.

Comparer cette courbe à la courbe théorique de la page 226 (en ayant soin d'adopter les mêmes échelles).

2^e^ Section

Cas des probabilités variables. — Supposons que des groupes d'épreuves répétées se produisent, dans lesquels un événement se présente avec des probabilités variant d'une épreuve à l'autre ; nous supposons cependant que, d'un groupe à l'autre, les probabilités se reproduisent dans le même ordre. Par exemple, chaque série d'épreuves consiste à tirer une boule succcessivement de cinq urnes données en remettant à chaque fois la boule dans l'urne dont on l'a tirée. Si les urnes sont de différentes compositions, les probabilités de tirer une blanche de chacune des urnes sont respectivement p_1, p_2, p_3, p_4, p_5. — Cherchons la valeur moyenne de la fréquence dans une série de n épreuves d'un événement fortuit E, connaissant la probabilité p_1 dans la pre-

mière, p_2 dans la deuxième, ... p_n dans la dernière.

Cette valeur moyenne est aussi l'espérance mathématique de recevoir une somme de f francs pour chacune des valeurs réalisées f de la fréquence de E dans les n épreuves. Mais au lieu de donner à la fin des épreuves $\frac{r}{n}$ francs si r épreuves ont été favorables, on peut donner $\frac{1}{n}$ francs pour chacune de celles des épreuves où s'est produit l'événement. Cette espérance est donc l'espérance mathématique de recevoir $\frac{1}{n}$ dans une épreuve en cas de réalisation d'un événement de probabilité p_1, $\frac{1}{n}$ pour p_2 : etc. ; c'est donc :

$$\frac{1}{n} p_1 + \frac{1}{n} p_2 + \ldots + p_n \frac{1}{n} = \frac{p_1 + p_2 + \ldots + p_n}{n};$$

c'est donc la moyenne arithmétique des n probabilités.

On peut encore raisonner ainsi : cherchons la valeur moyenne V′ du nombre de fois r que se réalise E dans les n épreuves. Ce nombre r est la somme $X_1 + X_2 + \ldots + X_n$ du nombre de fois que l'événement se réalise dans la 1^re^, la 2^e^, ... la $n^{\text{ième}}$ épreuve : X_1, X_2, ... ne pouvant donc prendre que les valeurs 0 ou 1. Alors V′ est la somme des valeurs moyennes de X_1, ... X_n. Mais la valeur moyenne de X_1 est comme on l'a vu page 148, la somme des produits de chacune des valeurs possibles (1 ou 0) de X_1, par la probabilité (p_1 ou $1 - p_1$) que X_1 prenne cette valeur. C'est donc

$1 \times p_1 + 0\,(1 - p_1) = p_1$, etc..... Donc $V' = p_1 + p_2 + \ldots\ldots + p_n$. Et la valeur moyenne, non pas V' de r, mais V de $f = \frac{r}{n}$ sera $\frac{V'}{n}$, soit :

$$V = \frac{p_1 + \ldots + p_n}{n}.$$

Cherchons maintenant l'écart quadratique moyen, λ, de r : ce sera l'écart quadratique moyen d'une somme $X_1 + \ldots\ldots + X_n$ de n nombres aléatoires indépendants, si nous supposons comme nous allons le faire maintenant que la réalisation de E dans chaque épreuve est indépendante de sa réalisation dans les autres. On peut alors utiliser la formule (21) de la page 160 ; ici

$$\lambda^2 = \lambda_1^2 + \ldots + \lambda_n^2.$$

λ_1 désigne l'écart quadratique moyen de X_1 ; donc son carré λ_1^2 est la valeur moyenne du carré de la différence de X_1 avec sa valeur moyenne p_1 ; comme il y a une probabilité p_1 que $X_1 - p_1 = 1 - p_1$ et une probabilité $1 - p_1$ que $X_1 - p_1 = 0 - p_1$, on a, par définition de la valeur moyenne :

$$\begin{aligned} \lambda_1^2 &= p_1\,(1 - p_1)^2 + (1 - p_1)\,(0 - p_1)^2 \\ &= p_1\,q_1^2 + q_1\,p_1^2 = p_1\,q_1\,(p_1 + q_1) = p_1\,q_1 \end{aligned}$$

en posant $q_1 = 1 - p_1$ Donc

$$\lambda^2 = p_1\,q_1 + \ldots + p_n\,q_n.$$

L'écart quadratique moyen μ de la fréquence $\frac{r}{n}$ sera $\frac{1}{n}$ fois celui λ de r, donc :

$$\mu = \frac{\sqrt{p_1 q_1 + \ldots + p_n q_n}}{n}.$$

On voit en particulier que, les quantités p_1, q_1, ... étant $\geqq 0$ et $\leqq 1$ on a

$$p_1 q_1 < 1, \ldots p_n q_n < 1, \text{ d'où :}$$

$$\lambda = \sqrt{p_1 q_1 + \ldots + p_n q_n} < \sqrt{n}, \ \mu < \frac{\sqrt{n}}{n}$$

$$\mu < \frac{1}{\sqrt{n}}.$$

On en conclut immédiatement *le théorème de Poisson* : la probabilité $P_n^{(\varepsilon)}$ pour que la fréquence d'un événement fortuit dans un groupe de n épreuves indépendantes diffère de plus d'un nombre arbitraire ε de la moyenne arithmétique p_0 des probabilités de cet événement dans ces n épreuves tend vers zéro lorsque le nombre n de ces épreuves croît indéfiniment.

En effet la probabilité P_t pour que $|f - p_0| > t\mu$ est $< \frac{1}{t^2}$. Prenons t de sorte que $t\mu = \varepsilon$; alors $P_n^{(\varepsilon)} = P_t < \frac{1}{t^2}$ ou $< \frac{\mu^2}{\varepsilon^2}$. Quand n croît indéfiniment, ε est fixe, $\mu^2 < \frac{1}{n}$ tend vers zéro. Donc $P_n^{(\varepsilon)} < \frac{\mu^2}{\varepsilon^2}$ tend vers zéro. On voit même que $P_n^{(\varepsilon)}$ est en langage mathématique au moins de l'ordre de $\frac{1}{n}$, c'est-à-dire que $\frac{P_n^{\varepsilon}}{\frac{1}{n}}$ est borné,

reste inférieur à un nombre indépendant de n, par exemple ici au nombre $\frac{1}{\varepsilon^2}$.

Ainsi, lorsqu'on mesurera la fréquence de E dans n épreuves, si n est assez grand et si on compare les fréquences relatives des groupes très nombreux de n épreuves, ces fréquences seront très probablement voisines d'un nombre p_0 qui est la moyenne arithmétique $\frac{p_1 + \dots + p_n}{n}$ des probabilités de E dans chaque épreuve. *A ce point de vue* tout se passera comme si E au lieu d'avoir une probabilité variable dans le groupe de n épreuves avait une probabilité fixe $p_0 = \frac{p_1 + \dots + p_n}{n}$.

Il y a cependant une différence entre les deux cas qu'on peut appeler cas de Bernoulli et cas de Poisson. Si le cas de Poisson se comportait exactement comme celui d'une probabilité constante $p_0 = \frac{p_1 + \dots + p_n}{n}$, l'écart quadratique moyen de la fréquence serait $\mu_0 = \sqrt{\frac{p_0 q_0}{n}}$ en posant $q_0 = 1 - p_0$, alors que cet écart est :

$$\mu = \sqrt{\frac{p_1 q_1 + \dots + p_n q_n}{n^2}}.$$

Donc

$$\mu_0^2 - \mu^2 = \frac{p_0 q_0}{n} - \frac{p_1 q_1 + \dots + p_n q_n}{n^2}$$

$$= \frac{np_0 q_0 - p_1 q_1 \dots - p_n q_n}{n^2} = \frac{(p_0 q_0 - p_1 q_1) + \dots + (p_0 q_0 - p_n q_n)}{n^2},$$

Or

$$p_0 q_0 - p_i q_i = (p_0 - p_i)(q_i - q_0) + 2 p_0 q_0 - p_i q_0 - p_0 q_i$$

et puisque

$$q_i - q_0 = (1 - p_i) - (1 - p_0) = (p_0 - p_i),$$
$$p_0 q_0 - p_i q_i = (p_0 - p_i)^2 + 2 p_0 q_0 - p_i q_0 - p_0 q_i.$$

Donc en remplaçant successivement i par $1, 2, \ldots . n$ et ajoutant

$$n^2 (\mu_0^2 - \mu^2) = \Sigma (p_0 - p_i)^2 + 2 n p_0 q_0 - \Sigma q_0 p_i - \Sigma p_0 q_i$$

et comme :

$$\Sigma q_0 p_i = q_0 p_1 + q_0 p_2 + \ldots = q_0 (p_1 + p_2 + \ldots) = n p_0 q_0$$
$$\Sigma p_0 q_i = p_0 q_1 + p_0 q_2 + \ldots = p_0 (q_1 + q_2 + \ldots) = n p_0 q_0$$

il reste

$$n^2 ({\mu_0}^2 - \mu^2) = \Sigma (p_0 - p_i)^2. \tag{36}$$

Le second nombre ne peut être négatif ; donc $\mu^2 \leqq {\mu_0}^2$. Et il ne peut être nul, c'est-à-dire, on ne peut avoir $\mu^2 = {\mu_0}^2$, que si $p_0 - p_1$, $p_0 - p_2, \ldots .$ sont nuls, c'est-à-dire si les probabilités $p_1, p_2, \ldots$ sont toutes égales, autrement dit si on n'est pas dans le cas de Poisson. Donc dans le cas de Poisson l'écart quadratique moyen μ est inférieur (et non pas égal) à ce qu'il serait si la valeur moyenne des fréquences, soit p_0, restant la même, on était dans le cas de Bernoulli.

Prenons par exemple $n = 2$ et

$$p_1 = \frac{1}{8} \cdot p_2 = \frac{7}{8} \text{ ; on aura } p_0 = \frac{1}{2}\left(\frac{1}{8} + \frac{7}{8}\right) = \frac{1}{2}$$

$$\mu = \frac{\sqrt{p_1 q_1 + p_2 q_2}}{n} = \frac{\sqrt{\frac{1}{8}\frac{7}{8} + \frac{7}{8}\frac{1}{8}}}{2} = \frac{\sqrt{14}}{16}$$

$$\mu_0 = \frac{\sqrt{p_0 q_0}}{n} = \frac{\sqrt{\frac{1}{2}\frac{1}{2}}}{2} = \frac{1}{2\sqrt{2}} = \frac{\sqrt{2}}{4} = \frac{4\sqrt{2}}{16}.$$

On voit que

$$\frac{\mu_0}{\mu} = \frac{4\sqrt{2}}{\sqrt{14}} = \sqrt{\frac{16}{7}} = 1{,}51.$$

Ainsi μ_0 est moitié plus grand que μ dans cet exemple particulier.

Loi des écarts dans le cas des probabilités variables. Cas de Poisson. — Poisson a pu démontrer que dans le cas des probabilités variables les écarts sont régis par la même loi que dans le cas des probabilités constantes :

Soit $P_n(t_1, t_2)$ la probabilité pour que l'écart $f - p_0$ entre la fréquence dans une série de n épreuves et la moyenne arithmétique $p_0 = \frac{p_1 + \ldots + p_n}{n}$ des probabilité soit comprise entre t_1 fois et t_2 fois l'écart quadratique moyen :

$$\mu_n = \frac{\sqrt{p_1 q_1 + \ldots + p_n q_n}}{n} \text{ de } f.$$

1° Cette probabilité a, lorsque n croît indéfiniment, une limite indépendante de l'événement fortuit et de la catégorie d'épreuves envisagée.

2° Cette limite est la différence $\psi_0(t_2) - \psi_0(t_1)$ des valeurs que prend une certaine fonction $\psi_0(t)$ dont il est possible de donner une table numérique.

L'énoncé implique que la fontion ψ_0 ne dépendant pas de $p_1, p_2, \ldots\ldots$ sera la même si on suppose $p_1 = p_2 = \ldots\ldots = p_n$, ce qui est le cas de Bernoulli ; il implique donc que la fonction ψ_0 est celle que nous avons déjà rencontrée, page 219, sous le nom de ψ.

On prend généralement $t_1 = -t_2 = -t$. Alors P_n est la probabilité pour que

$$-t\mu_n < f - p_0 < t\mu_n \quad \text{ou} \quad |f - p_0| < t\mu_n.$$

Alors $P_n(t, -t)$ tend vers $\psi_0(t) - \psi_0(-t)$, et comme la fonction ψ_0 est paire, c'est-à-dire $\psi_0(-t) = -\psi_0(t)$ quel que soit t, on a

$$\psi_0(t) - \psi_0(-t) = 2\psi_0(t)$$

En appelant comme précédemment $\Psi(t)$ cette fonction, on voit que $\Psi(t)$ est *la limite de la probabilité* pour que $|f - p_0| < t\mu_n$, lorsque n croît indéfiniment.

En appliquant ce théorème de Poisson, il faut bien prendre garde que μ_n n'a pas pour valeur l'expression $\sqrt{\dfrac{p_0 q_0}{n}}$ qu'on obtiendrait si les valeurs des probabilités, au lieu d'avoir simplement p_0 comme moyenne arithmétique dans les n épreuves, étaient égales à p_0 dans chacune d'elles. Il faut prendre la valeur plus petite :

$$\mu_n = \frac{\sqrt{p_1 q_1 + \ldots + p_n q_n}}{n},$$

Pour bien nous rendre compte de la différence des résultats, reprenons l'exemple utilisé précédemment :

Supposons qu'on ait cent une (101) urnes contenant chacune cent boules blanches ou noires, la première contenant o boule blanche, la seconde, une, la cent-unième, rien que des blanches.

Les probabilités correspondantes d'extraire une blanche seront :

$$p_1 = 0\,;\; p_2 = \frac{1}{100}\,;\ldots p_{101} = \frac{100}{100}\,.$$

La moyenne arithmétique $p_0 = \dfrac{p_1 + p_2 + \ldots\ldots + p_{101}}{101}$ $= \dfrac{1}{2}$ séra aussi la probabilité de tirer une blanche d'une urne contenant autant de blanches que de noires, par exemple d'une cent deuxième urne où on aurait versé le contenu des cent une premières.

Si on effectue un grand nombre N de groupes de 101 tirages des 101 premières urnes, en remettant les boules immédiatement après chaque tirage, la valeur de la fréquence pour chaque groupe ne pourra être que o, $\dfrac{1}{101}$, $\dfrac{2}{101}$, ... ou 1, et la moyenne de ces N fréquences sera voisine de $p_0 = \dfrac{1}{2}$, si N est assez grand.

Si on fait un grand nombre N de séries de cent un tirages dans la cent deuxième urne, la fréquence dans chaque série ne peut encore prendre que les valeurs o, $\dfrac{1}{101}$, $\dfrac{2}{101}$, ... ou 1, et la moyenne de ces N fréquences sera encore voisine de $p_0 = \dfrac{1}{2}$.

Mais cherchons à évaluer quelle proportion $\frac{m}{N}$ des N séries donnera une fréquence f différant de $\frac{1}{2}$ de moins de $\frac{3}{100}$ par exemple. Posons :

$$t\,\mu_n = \frac{3}{100}, \text{ d'où } t = \frac{0{,}03}{\mu_n}.$$

On aura approximativement :

$$\frac{m}{N} = \Psi(t) = \Psi\left(\frac{0{,}03}{\mu_n}\right)$$

S'il s'agit de tirages dans la cent deuxième urne, on aura pour μ_n la valeur

$$\mu'_n = \sqrt{\frac{p_0\,q_0}{n}} = \sqrt{\frac{\frac{1}{2}\,\frac{1}{2}}{101}} = \frac{1}{2\sqrt{101}}$$

S'il s'agit de groupes de 101 tirages dans les 101 premières urnes :

$$n^2\,\mu^2_n = p_1\,q_1 + p_2\,q_2 + \ldots = 0 \times 1 + \frac{1}{100} \times \frac{99}{100}$$
$$+ \frac{2}{100} \times \frac{98}{100} + \ldots + \frac{99}{100} \times \frac{1}{100} + 1 \times 0\,;$$
$$100^2 \times 101^2\,\mu^2_n = 1 \times 99 + 2 \times 98 + 3 \times 97 + \ldots + 50 \times 50$$
$$+ 51 \times 49 + \ldots + 99 \times 1 =$$
$$[1 \times 99 + 2 \times 98 + \ldots + 49 \times 51] \times 2 + 50 \times 50 = 2S + 50^2$$

en appelant S la quantité entre crochets.

On peut calculer le second terme par des procédés élémentaires ou le calculer directement ce qui ne serait qu'un peu plus long.

Par exemple on peut écrire : le r^e terme entre crochets : $r\,(100 - r) = 100\,r - r^2$.

Donc : $S = \Sigma\, 100\, r - \Sigma\, r^2 =$ 100 fois la somme des 49 premiers nombres entiers successifs — la somme des carrés des 49 premiers nombres entiers successifs $= 100\,(1 + 2 + \ldots\ldots + 49) - (1^2 + 2^2 + \ldots\ldots + 49^2)$.

Or, en arithmétique élémentaire on montre que $1 + 2 + \ldots\ldots + q = \frac{q\,(q+1)}{2}$, et que $1^2 + 2^2 + \ldots\ldots + q^2 = \frac{q\,(q+1)\,(2\,q+1)}{6}$, quel que soit l'entier q. On a donc ici :

$$1 + 2 + \ldots + 49 = \frac{49 \times 50}{2} = 25 \times 49$$

$$1^2 + 2^2 + \ldots + 49^2 = \frac{49 \times 50 \times (49 \times 2 + 1)}{6}$$

$$= \frac{49 \times 50 \times 99}{6} = 49 \times 25 \times 33.$$

D'où

$$S = 100 \times 25 \times 49 - 49 \times 25 \times 33 = 25 \times 49 \times 67.$$

Alors

$$\mu^2_n = \frac{2\,S + 50^2}{101^2\ 100^2} = \frac{50\,[49 \times 67 + 50]}{101^2\ 100^2} = \frac{50 \times 3.333}{101^2\ 100^2}$$

$$= \frac{166.650}{(10.100)^2}$$

$$\mu_n = \frac{\sqrt{166.650}}{10.100} = \frac{408}{10.100} = 0{,}0404, \text{ tandis que } \mu'_n$$

$$= \frac{1}{2\sqrt{101}} = \text{environ } 0{,}05.$$

On a bien $\mu_n < \mu_n$. — Alors, dans le cas des cent une premières urnes :

$$t = \frac{0{,}03}{\mu_n} = \frac{0{,}03}{0{,}0404} = 0{,}743$$

et dans le cas de la dernière urne :

$$t = \frac{0{,}03}{0{,}05} = \frac{3}{5} = 0{,}6.$$

Alors 1° cas :

$$\frac{m}{N} = \Psi\,(0{,}743) = \text{environ } 0{,}54$$

2° cas :

$$\frac{m}{N} = \Psi\,(0{,}6) = \text{environ } 0{,}45.$$

En définitive, si un observateur voit successivement effectuer 100 séries de 101 tirages, dans le cas de Poisson et dans le cas de Bernoulli (urnes distinctes, et urne unique), il constatera que dans les deux cas les valeurs de f se groupent autour de 0,5. Mais dans le cas de Poisson il y en aura environ 54 où la fréquence différera de $\frac{1}{2}$ de moins de $3/100$, il n'y en aura que 45 environ dans le cas de Bernoulli. Le groupement autour de 0,5 sera donc plus serré dans le cas de Poisson. Ce résultat, très intéressant, permet d'éviter des erreurs qu'on aurait pu commettre, si on n'avait pas connu le cas de Poisson. Si le groupement des fréquences autour d'une valeur p_0 est plus serré que ne l'indique la formule

de Laplace, il sera bon de se demander si la probabilité qu'on avait crue constante ne serait pas variable à l'intérieur de chaque groupe d'épreuves.

EXERCICE

(F) Deux urnes A, B contiennent chacune une boule blanche et une boule noire, identiques à la couleur près.

On effectue des groupes de deux tirages, un dans chaque urne. On demande de calculer la valeur moyenne du nombre r de boules blanches sorties dans chaque groupe.

Montrer que cette valeur moyenne reste la même si on met avant les tirages les deux boules blanches dans A, les deux noires dans B.

Montrer qu'au contraire les écarts quadratiques moyens de r correspondants sont différents.

Examen des statistiques.

Laissons pour le moment de côté les statistiques qui donnent les répétitions respectives des différentes valeurs d'une grandeur variable telle que la taille humaine, etc. Considérons celles qui donnent les répétitions d'un même événement dans des groupes d'épreuves en nombre égal ou à peu près égal. Par exemple la statistique hebdomadaire du nombre des décès (et, de même, du nombre des naissances, etc.) à Paris.

Supposons qu'elles se groupent autour d'une valeur centrale : on serait conduit à se demander si cet événe-

ment obéit à la loi du hasard, c'est-à-dire s'il a une probabilité déterminée. Pour l'établir il faut, pour un nombre d'observations assez grand, que les répétitions soient assez voisines. Mais c'est là une notion vague, et qu'il importe de préciser. — Un premier moyen consiste à dresser le tableau des répétitions des différents écarts avec la valeur qu'on suppose être la probabilité, et à comparer ce tableau avec celui de $\psi(t)$.

Il n'est pas nécessaire de représenter graphiquement les écarts eux-mêmes (ce qui obligerait à déterminer d'abord la probabilité qu'on cherche); on relèvera les répétitions R_i des valeurs observées de la fréquence f comprises dans un petit intervalle ε_i, ε_{i+1}. On tracera un rectangle d'aire R_i ayant pour base l'intervalle de ε_i à ε_{i+1}. On opèrera de même pour une suite d'intervalles ε_i..... et on obtiendra une suite de rectangles accolés qui déterminent une ligne polygonale. La forme de la courbe permet tout de suite d'éliminer les cas où les valeurs de la fréquence ne se groupent pas autour d'une valeur centrale comme le veut la formule de Laplace (page 207) (si elles se groupent suivant cette formule, on dira que la distribution des écarts est typique). Généralement la ligne polygonale obtenue ne rappelle que d'assez loin la courbe en cloche. Pour aller plus loin, on utilisera la courbe en ogive.

On considèrera le nombre S_x des séries où la fréquence d'un événement E est inférieure à x.

Pour chaque valeur de x, on reporte sur le graphique le nombre S_x. Si on obtient une courbe d'une forme

assez analogue à celle que nous avons rencontrée (page 221) et appelée courbe en ogive, on poursuit la comparaison en construisant *le graphique réduit*.

Si je suppose que $f_1 = \frac{m_1}{n}$, où m_1 est le nombre des n épreuves de la première série qui sont favorables, $f_2 = \frac{m_2}{n}$, etc., j'ai : $\frac{m_1 + m_2 + \ldots + m_N}{N} = f_0 =$ la fréquence de E dans l'ensemble de toutes les épreuves. Si elles sont assez nombreuses, ce quotient sera une valeur approchée de la probabilité p (à supposer qu'elle existe) de l'événement E. Le carré de l'écart quadratique moyen de f avec f_0 a comme valeur approchée (voir page 168) :

$$\frac{(f_0 - f_1)^2 + (f_0 - f_2)^2 + \ldots + (f_0 - f_N)^2}{N - 1}. \tag{39}$$

Pour former la courbe réduite, nous portons comme abscisse non pas la fréquence x, mais la fréquence réduite : $t = \frac{x}{\mu}$. D'autre part nous portons en ordonnée non pas le nombre total des répétitions, mais la fréquence correspondante $y = \frac{S_x}{N}$. — Si la courbe réduite ainsi obtenue réalise une coïncidence suffisante avec la courbe théorique en ogive telle qu'elle résulte de la formule de Laplace on dira qu'elle est *typique*. C'est la première condition de l'existence d'une probabilité.

Mais cette condition n'est pas suffisante, et c'est ce qui ressort de l'exemple de Poisson. En effet, dans le

cas des urnes distinctes, la distribution des écarts est typique, et cependant les différentes épreuves de chaque groupe n'ont pas les mêmes probabilités, bien que ces diverses probabilités p_1, ... p_n se retrouvent les mêmes à chaque groupe. Il y a donc une autre condition.

Le premier qui l'ait précisée est l'actuaire français Dormoy (coefficient de divergence).

Le statisticien allemand Lexis a employé, pour vérifier sa réalisation, une méthode analogue : c'est le calcul du *coefficient de dispersion*. — Lorsque nous avons calculé l'écart quadratique moyen μ, nous avons employé une formule approchée (39), déduite d'une définition de l'écart quadratique moyen qui s'applique même à des événements qui n'ont pas une probabilité déterminée. Au contraire, dans les cas Bernoulli-Poisson, nous avons déduit de la définition de μ les valeurs qui peuvent le remplacer quand on sait qu'on est bien dans le cas de Poisson ou de Bernoulli. Par exemple, dans le cas de Bernoulli, on a $\mu = \sqrt{\frac{p_0 q_0}{n}}$; dans le cas de Poisson (séries d'épreuves de probabilité p_1 pour la 1^{re} épreuve de chaque série, p_2 pour la 2^e, etc.) on a :

$$\mu = \frac{\sqrt{p_1 q_1 + p_2 q_2 + \cdots + p_n q_n}}{n}.$$

La seconde valeur est plus petite que la première où on appellerait p_0 la valeur moyenne des fréquences, p_1, etc.

Nous supposons qu'on a constaté que les valeurs des

fréquences sont pour la plupart voisines de f_0. Si on est dans le cas de Bernoulli, f_0 est voisin de p_0, et q_0 de $1 - f_0$. Si on est dans le cas de Poisson, ce sera encore vrai, puisque $p_0 =$ la valeur moyenne des fréquences. Donc $\sqrt{\frac{p_0 q_0}{n}}$ est voisin de $\sqrt{\frac{f_0(1-f_0)}{n}}$ Nous pouvons appeler la première de ces deux quantités, à savoir

$$\mu_0 = \sqrt{\frac{(f_1 - f_0)^2 + (f_2 - f_0)^2 + \dots}{N - 1}}.$$

l'écart quadratique moyen physique ou empirique, et la seconde,

$$\mu'_0 = \sqrt{\frac{f_0(1 - f_0)}{n}}.$$

l'écart quadratique moyen théorique. — Lexis a appelé *coefficient de dispersion* le rapport de ces deux quantités, $\frac{\mu_0}{\mu_0'}$, c'est-à-dire :

$$(10) \qquad \frac{\sqrt{\frac{(f_1 - f_0)^2 + \dots\dots}{N - 1}}}{\sqrt{\frac{f_0(1 - f_0)}{n}}} = Q.$$

1° Si on est dans le cas de Bernoulli, c'est-à-dire si l'événement considéré a une probabilité constante p_0. Q sera en général très voisin de 1.

2° Si on est dans le cas de Poisson, c'est-à-dire si,

d'une série à l'autre, les épreuves gardent les mêmes probabilités, ces probabilités variant dans chaque série, on aura $Q < 1$.

3° La quantité Q peut être calculée dans tous les cas sans faire aucune hypothèse, sans savoir si les groupes considérés se rapportent à des événements qui aient aucun rapport les uns avec les autres : elle peut l'être, à plus forte raison, si ce sont des événements sociaux du même genre.

Nous dirons que, par définition, la distribution des fréquences des écarts est *normale*, si $Q = 1$ sensiblement. Si, comme dans le cas de Poisson, $Q < 1$, cette distribution est dite *hyponormale*.

La valeur de Q, calculée dans la plupart des tableaux statistiques, n'est ni égale à 1, ni inférieure à 1, mais en général > 1. On a alors une distribution des fréquences ou des écarts qui est *hypernormale*.

Avant Lexis, Dormoy avait procédé de la même façon en partant de l'écart moyen (et non de l'écart quadratique moyen). On forme l'écart moyen physique :

$$\sigma_0 = \frac{1}{N}\left[\,|f_1 - f_0| + |f_2 - f_0| + \ldots\right]$$

et la valeur $\sigma'_0 = \sqrt{\frac{2}{\pi}}\sqrt{\frac{f_0(1-f_0)}{n}}$, valeur approchée de l'écart moyen théorique dans le cas de Bernoulli (page 174) : Et on forme le *coefficient de divergence* $\frac{\sigma_0}{\sigma'_0} = S$. Si on est dans le cas de Bernoulli, S sera

voisin de 1. Le procédé de Dormoy est non seulement antérieur à celui de Lexis, mais il conduit à des calculs numériques plus simples, la quantité σ_0 se calculant beaucoup plus rapidement que μ_0. Il serait à souhaiter qu'on l'employât de préférence[1].

— Il y a des séries statistiques qui vérifient les deux conditions indiquées, c'est-à-dire telles que la distribution des écarts ou des fréquences y soit à la fois typique et normale.

Par exemple[2], la fréquence annuelle f des garçons parmi les nouveau-nés dans 26 années consécutives a été :

0,5164	0,5155	0,5152
0,5166	0,5161	0,5145
0,5150	0,5140	0,5138
0,5157	0,5150	0,5144
0,5151	0,5149	0,5157
0,5156	0,5140	0,5132
0,5159	0,5145	0,5146
	0,5148	0,5151
	0,5147	0,5150
		0,5147

On voit : 1° qu'il est né à peu près autant de filles que de garçons ; 2° qu'il y a eu une légère prédominance

1. Il n'a pour nous que l'inconvénient de supposer connue la formule $\frac{\sigma}{\mu} = \sqrt{\frac{2}{\pi}}$ dans le cas de Bernoulli, formule dont la démonstration entraîne l'emploi du calcul intégral dont nous tenons ici à nous passer.

2. L'exemple suivant, ainsi que l'exercice page 244, ont été proposés, en 1907, à l'examen des statisticiens.

des garçons ; 3° que les écarts des 26 valeurs de f, à partir de la moyenne $f_0 = 0{,}5150$, sont assez petits. On est conduit à se demander si la naissance d'un garçon est un événement qui a une probabilité déterminée, soumise aux lois du hasard.

Pour le savoir, on adopte provisoirement comme valeur de cette probabilité $p = f_0 = 0{,}5150$. Alors, si l'hypothèse est exacte, l'écart quadratique moyen théorique est :

$$\mu'_0 = \sqrt{\frac{p(1-p)}{n}}, \text{ où } p = 0{,}515, \text{ et } n = 734.449$$

(nombre moyen annuel de naissances). D'où : $\mu'_0 = 0{,}000583$. D'autre part l'écart quadratique moyen empirique μ_0 est donné par :

$$\mu_0 = \sqrt{\frac{(f_1 - f_0)^2 + \dots}{N - 1}}$$

où N (nombre des séries) $= 26$;

$f_1 - f_0 = 0{,}5164 - 0{,}5150 = 0{,}0014$; $f_2 - f_0 = \dots\dots$ On trouve :

$$\mu_0 = \sqrt{\frac{0{,}00001612}{25}} = 0{,}000803.$$

Le coefficient de dispersion $Q = \frac{\mu_0}{\mu'_0}$ peut alors se calculer, et l'on trouve : $Q = 1{,}37$. — On voit que la série est hypernormale, mais assez peu.

Elle est presque normale (la quantité $Q^2 - 1 = 0{,}88$).

Elle le serait peut-être entièrement, d'après M. Carvallo, si les déclarations des embryons n'étaient point parfois inexactes.

Cherchons maintenant si la distribution des écarts est typique, c'est-à-dire conforme à la formule de Laplace. Nous calculerons les écarts réduits $t = \frac{\varepsilon}{\mu_0}$. Puisqu'on n'a tenu compte, pour les f, que des $\frac{1}{10000}$, on prendra pour les ε des valeurs intermédiaires, par exemple $\varepsilon = 0{,}00025$; $\varepsilon = 0{,}00045$; ..... ; en regardant la table, on cherchera quel est le nombre R des valeurs de f compris entre $f_0 \pm \varepsilon = 0{,}515 \pm \varepsilon$; on obtiendra la colonne des R observés.

Puis on calculera théoriquement le même nombre, en remarquant que $\frac{R}{N}$ est environ la probabilité d'un écart réduit en valeur absolue $\leq t$, c'est-à-dire $\Psi(t) = \Psi\left(\frac{\varepsilon}{\mu_0}\right)$. On mettra alors dans la colonne R calculé le produit N $\Psi(t)$. On aura ainsi :

100 000 ε	25	45	65	85	105	125	145	165	185
R observés	8	11	16	18	21	23	24	25	26
R calculés	6,4	11,1	15,3	18,5	21,1	22,9	24,2	24,9	25,4

On voit qu'on obtient une concordance bien meilleure que Carvallo (page 45) qui prend comme écart réduit $\frac{\varepsilon}{\mu'_0}$, et qui place ε exactement au bout des intervalles 2, 4, 6,.....

Les résultats (observés et calculés) sont tous égaux à une unité près (sauf le premier, où la différence est 1 1/2). Ainsi on a une distribution sensiblement typique et normale qui a tous les caractères principaux qu'offrirait une distribution de fréquences relatives à un événement fortuit de probabilité constante $p = 0{,}515$.

Nous allons maintenant donner un exemple d'une série dont la distribution est hypernormale d'une façon beaucoup plus marquée.

Utilisons, comme Arne Fisher (page 153) la statistique qui donne le nombre annuel des mariages en Danemark. On a réduit la population supposée stationnaire à 2.500.000.

1888.	17.605	1901	17.870
1889.	17.622	1902	17.712
1890.	17.181	1903	17.791
1891.	17.017	1904	17.895
1892.	17.012	1905	17.947
1893.	17.676	1906	18.592
1894.	17.445	1907	19.072
1895.	17.736	1908	18.750
1896. . . .	18.239	1909	18.453
1897.	18.676	0910	18.225
1898.	18.870	1911	17.749
1899.	18.661	1912	18.034
1900.	19.015		

On trouve que la moyenne de ces 25 nombres est 18.035, et que leur écart quadratique moyen est 588,43. La fréquence moyenne est donc $f_0 = \dfrac{18.035}{2.500.000}$ dans

une suite de $N = 25$ série d'observations, chaque série portant sur $n = 2.500.000$ épreuves.

L'écart de ces fréquences est $\mu_0 = \dfrac{588,43}{2.500.000}$. Si donc l'événement : mariage d'un individu déterminé dans l'année, pouvait être considéré comme un événement fortuit, sa probabilité serait p_0, voisine de f_0, et son écart quadratique moyen serait voisin de

$$\mu'_0 = \sqrt{\frac{f_0(1 - f_0)}{n}}.$$

On trouve que

$$\mu'_0 = \frac{133,81}{2.500.000}$$

de sorte que le coefficient de dispersion

$$Q = \frac{\mu_0}{\mu'_0} = \frac{588,43}{133,81} = 4,41.$$

Ainsi on a $Q > 1$; et $Q^2 - 1 = 4,41^2 - 1 = 19,45 - 1 = 18,45$. On a une série nettement hypernormale. On ne peut donc considérer la probabilité annuelle des mariages comme constante.

L'étude du cas de Poisson avait déjà permis d'établir que des valeurs de Q différentes de 1 n'étaient pas incompatibles avec la théorie des probabilités. Il restait à imaginer un cas simple où cette théorie permettrait de prévoir des valeurs de $Q > 1$. Tels sont ceux qu'ont étudiés Lexis et Borel.

3° Section

Le cas de Lexis. — Nous avons vu que, dans le cas de Poisson, on compare les fréquences f_1, f_2,..... f_N, obtenues dans des groupes de n épreuves, la probabilité variant dans un même groupe d'une épreuve à l'autre, mais les n valeurs de ces probabilités se reproduisant d'un groupe à l'autre. Dans le cas de Lexis, c'est l'inverse qu'on suppose. La probabilité est constante d'une épreuve à l'autre dans un même groupe, mais elle varie d'un groupe à l'autre. C'est, par exemple, ce qui se produirait si, pendant les épreuves du premier groupe, on tirait successivement n boules (en remettant chaque boule une fois tirée après avoir noté sa couleur) d'une première urne, puis, de même, pendant la deuxième série, d'une urne de composition différente, etc.

Soient donc p_1 la probabilité constante dans les n épreuves du 1er groupe, p_2 la probabilité constante dans les n épreuves du 2^e groupe,..... p_N la probabilité constante dans les n épreuves du N^e groupe

Notons d'autre part les fréquences $f_1 = \frac{m_1}{n}$,..... $f_N = \frac{m_N}{n}$ de l'événement favorable dans les 1er,..... N^e groupes.

Nous allons calculer la valeur moyenne et l'écart quadratique moyen des N fréquences f_1,..... f_N.

La moyenne arithmétique de ces N fréquences est

$$f_0 = \frac{f_1 + \ldots + f_N}{N} = \frac{\frac{m_1}{n} + \ldots + \frac{m_N}{n}}{N} = \frac{m_1 + \ldots + m_N}{nN}.$$

C'est donc aussi la fréquence de l'événement fondamental dans le groupe des Nn épreuves formé par les N groupes des n épreuves.

Il s'agit maintenant de chercher à évaluer f_0, connaissant $p_1 \ldots . p_N$.

Si n est assez grand, $f_1 \ldots . f_N$ seront généralement voisins des probabilités respectives $p_1 \ldots . p_N$.

Donc, si n est suffisamment grand, la moyenne arithmétique f_0 de $f_1 \ldots . f_N$ sera également voisine de la moyenne arithmétique des probabilités des différentes séries : f_0 voisin de

$$p_0 = \frac{p_1 + \ldots + p_N}{N}$$

Remarquons en passant que si

$$|f_1 - p_1| < \varepsilon, \ldots |f_N - p_N| < \varepsilon,$$

comme on a

$$f_0 - p_0 = \frac{(f_1 - p_1) + \ldots + (f_N - p_N)}{N}$$

on a aussi

$$|f_0 - p_0| < \frac{\varepsilon + \ldots + \varepsilon}{N} = \varepsilon.$$

Par conséquent, ce qui est important pour que f_0 soit voisin de p_0, ce n'est pas la valeur de N, c'est que n soit grand.

Passons à l'écart quadratique moyen μ_0 des $f_1, f_2, \ldots\ldots$ avec p_0. On a

$$N\mu^2_0 = (f_1 - p_0)^2 + \ldots + (f_N - p_0)^2,$$

μ_0 étant voisin, si n est grand, de θ défini par

$$N\theta^2 = (p_1 - p_0)^2 + \ldots + (p_N - p_0)^2.$$

Nous pouvons préciser un peu. Pour cela nous allons imaginer qu'on répète indéfiniment N séries de n épreuves avec les mêmes probabilités p_1 dans la première série, p_2 dans la deuxième,..... p_N dans la N^e.

Appelons alors Mx la valeur moyenne de toute quantité x calculée successivement dans chaque groupe possible de N telles séries. On aura, pour chaque groupe de N séries, N valeurs de $f_1 \ldots\ldots f_N$, soient $f'_1 \ldots\ldots f'_N$; mais $p_1 \ldots\ldots p_N$ restent les mêmes.

Si on pose $f'_0 = \dfrac{f'_1 + \ldots\ldots + f'_N}{N}$, on remarque que, dans la catégorie des premières séries de chaque groupe (ou, si l'on veut, des tirages de la première urne), $Mf'_1 = p_1$, de même $Mf'_2 = p_2$, ... donc f'_0 a pour valeur moyenne $Mf'_0 = \dfrac{p_1 + \ldots\ldots + p_N}{N}$. Appelons p_0 cette quantité. Cherchons l'écart quadratique moyen θ de $f'_1, f'_2, \ldots\ldots f'_N$ avec p_0; $N\theta^2$ est la valeur moyenne de $(f'_1 - p_0)^2 + \ldots\ldots + (f'_N - p_0)^2$ et nous prendrons θ pour valeur approchée de μ_0.

Calculons par exemple la valeur moyenne de $(f'_1 - p_0)$. Nous avons :

$$f'_1 - p_0 = (f'_1 - p_1) + (p_1 - p_0),$$
$$(f'_1 - p_0)^2 = (f'_1 - p_1)^2 + (p_1 - p_1)^2 + 2(f'_1 - p_1)(p_1 - p_0).$$

La valeur moyenne de $(f'_1 - p_0)^2$ est donc égale à la somme de :

1° La valeur moyenne de $(f'_1 - p_1)^2$; c'est, comme nous l'avons vu dans le cas de Bernouilli : $\frac{p_1 q_1}{n}$.

2° La valeur moyenne de $(p_1 - p_0)^2$: cette quantité, étant fixe, est égale à sa valeur moyenne.

3° La valeur moyenne de $2(f'_1 - p_1)(p_1 - p_0)$. Or $2(p_1 - p_0)$ est fixe ; $f'_1 - p_1$ varie suivant la loi du hasard, mais sa valeur moyenne $M(f'_1 - p_1)$ est nulle.

Donc, en résumé, la valeur moyenne de $(f'_1 - p_0)^2$ est :

$$M(f'_1 - p_0)^2 = \frac{p_1 q_1}{n} + (p_1 - p_0)^2$$

et par suite la valeur moyenne $N\theta^2$ de

$(f'_1 - p_0)^2 + \ldots + (f'_N - p_0)^2$ sera $\frac{p_1 q_1}{n} + \ldots + \frac{p_N q_N}{n} + (p_1 - p_0)^2 + \ldots + (p_N - p_0)^2$.

Si n est grand, μ_0^2 sera donc voisin de cette quantité divisée par N. — Or on a vu, au sujet du cas de Poisson, que si l'on pose :

$$p_0 = \frac{p_1 + \ldots + p_n}{n}$$

on a (page 250) :

$$\frac{p_0 q_0}{n} - \frac{p_1 q_1 + \ldots + p_n q_n}{n^2} = \frac{1}{n^2}[(p_0 - p_1)^2 + \ldots + (p_0 - p_n)^2].$$

Ici on a posé

$$p_0 = \frac{p_1 + \ldots + p_N}{N};$$

on aura donc

$$\frac{p_0 q_0}{N} - \frac{p_1 q_1 + \dots + p_N q_N}{N^2} = \frac{(p_0 - p_1)^2 + \dots + (p_0 - p_N)^2}{N^2},$$

ou

$$p_1 q_1 + \dots + p_N q_N = \\ N p_0 q_0 - [(p_0 - p_1)^2 + \dots + (p_0 - p_N)^2].$$

D'où

$$\frac{p_1 q_1}{n} + \dots + \frac{p_N q_N}{n} = \frac{N}{n} p_0 q_0 - \frac{1}{n} [(p_0 - p_1)^2 + \dots \\ + (p_0 - p_N)^2].$$

Donc

$$N\theta^2 = \frac{N p_0 q_0}{n} + \left(1 - \frac{1}{n}\right) [(p_0 - p_1)^2 + \dots]$$

Mais observons que, si n est grand, μ_0 est voisin de sa valeur moyenne θ, et on pourra écrire :

$$\mu_0^2 = \frac{p_0 q_0}{n} + \left(1 - \frac{1}{n}\right) \frac{(p_0 - p_1)^2 + \dots + (p_0 - p_N)^2}{N}$$

Or nous avons vu que, si n est grand, la fréquence constatée dans les Nn épreuves est voisine de p_0. Si les probabilités $p_1 \dots p_N$ étaient égales entre elles et par conséquent à leur moyenne arithmétique p_0, l'écart quadratique moyen μ_0^2 serait voisin de $\mu'^2_0 = \frac{p_0 q_0}{n}$.

On a donc : $\mu_0^2 \geqq \mu'^2_0$ et on aura sûrement $\mu_0^2 > \mu'^2_0$ si les $p_1 \dots p_N$ ne sont pas égaux entre eux. Donc le coefficient de dispersion :

$$Q = \frac{\mu_0}{\mu'_0} = \tag{41}$$

$$\sqrt{1 + \frac{n-1}{N} \frac{(p_0 - p_1)^2 + \dots + (p_0 - p_N)^2}{p_0 q_0}} > 1$$

et ce coefficient sera d'autant plus grand que 1 que les probabilités $p_1 \ldots\ldots p_N$ seront plus différentes entre elles. Il ne serait égal à 1 que si ces probabilités étaient égales, c'est-à-dire si on se trouvait dans le cas de Bernoulli. Il lui est supérieur dans la mesure où on s'en écarte.

On peut préciser. Remarquons que :

$$Q^2 - 1 = \frac{n-1}{p_0 q_0} \times \mu''^2 = \frac{n-1}{n} \frac{\mu''^2}{\mu'^2_0}$$

où μ'' est l'écart quadratique moyen des probabilités p.

$$\left(\mu''^2 = \frac{(p_0 - p_1)^2 + \ldots + (p_0 - p_N)^2}{N}\right).$$

On voit donc que la quantité :

$$\mu'^2_0 (Q^2 - 1) \quad \text{ou} \quad \mu_0^2 - \mu'^2_0,$$

que nous savons calculer pour toute série statistique sans savoir si elle rentre ou non dans le cas de Lexis, est, dans le cas de Lexis, égale à $\left(1 - \frac{1}{n}\right) \mu''^2$ ou, si n est assez grand, à μ''^2.

On peut estimer que $\sqrt{\mu_0^2 - \mu'^2_0}$ mesure *le degré d'instabilité* de la probabilité d'une série à l'autre. Mais il le mesure pour une moyenne p_0 donnée. Pour mesurer la variation relative de p_0, soit $\frac{\text{écart } \mu'' \text{ de } p_0}{p^0}$, on considérera le rapport :

$$\frac{\sqrt{\mu_0^2 - \mu'^2_0}}{f_0},$$

qui peut être mesuré sans qu'on sache si la série est de Lexis ou non. On pourra le considérer, avec le statisticien suédois Charlier, comme le coefficient d'instabilité du groupe de séries. En réalité, comme c'est généralement une fraction décimale, Charlier prend ce nombre multiplié par 100.

Naturellement ceci n'a de sens que pour une distribution hypernormale.

Par exemple, s'il s'agit des naissances masculines, on trouve :

$$\frac{\sqrt{\mu_0^2 - \mu'^2_0}}{f_0} = \frac{\sqrt{(0{,}000803)^2 - (0{,}000583)^2}}{0{,}515} = 0{,}000107$$

On voit que l'écart quadratique moyen des probabilités de naissances masculines ne serait que de l'ordre du $\frac{1}{1.000}$ de cette probabilité.

Pour les mariages, on trouve :

$$\frac{\sqrt{\mu^2_0 - \mu'^2_0}}{f_0} = \frac{\sqrt{n^2\,\mu_0^2 - n^2\,\mu'^2_0}}{nf_0} = \frac{\sqrt{(588{,}43)^2 - (133{,}81)^2}}{18.035}$$
$$= \frac{572}{18.035} = 0{,}0316.$$

Ici, l'écart quadratique moyen des probabilités d'une série à l'autre serait les $\frac{3}{100}$ de cette probabilité. Il serait beaucoup plus sensible.

4e Section

Explication de l'origine des séries hypernormales par la dépendance de certains événements. Schéma des urnes de Borel. — Si on calcule la fréquence annuelle des décès, on s'aperçoit que, pour des individus du même âge, on obtiendrait des fréquences diverses si l'on groupait les citadins d'une part, les campagnards de l'autre, si l'on tenait compte de la profession, de la longévité des parents, etc. Bertrand a remarqué que ceci revient à supposer qu'on a fait des tirages dans des urnes de composition différente dans un même groupe, les urnes présentant d'ailleurs, d'un groupe à l'autre, la même variété de composition. C'est le cas de Poisson. Il ajoute que les causes de décès sont soumises à bien des vicissitudes : « C'est la disette qui rend les maladies plus abondantes et plus graves, la guerre qui accroît les chances ». Ceci revient à supposer qu'on a fait des tirages dans une urne dont la composition change d'un groupe de tirages à l'autre [1]. C'est le cas de Lexis. — Mais il ajoute qu'il y a une autre raison qui explique que la distribution des fréquences s'écarte de la loi normale. Considérons, par exemple, la fréquence des jours de pluie. « Quand il fait mauvais temps, ce n'est pas pour un jour seulement. » Ceci revient à supposer que la probabilité d'un événement est influencée

1. C'est-à-dire, dans le cas des décès, d'une année à l'autre.

par l'événement qui s'est produit avant celui-là. C'est le cas qui a été étudié par M. Borel.

M. Borel précise cette hypothèse à propos des décès. Certaines causes de décès agissent, en effet, de telle façon qu'un décès qui provient de telles causes provoque d'autres décès. Tels sont ceux qui sont dus à des maladies contagieuses. Si, par exemple, dans un groupe, un homme meurt du choléra, la probabilité de mourir pour ceux qui peuvent avoir avec lui des contacts directs et indirects est accrue. Il s'agit d'imaginer une combinaison où les probabilités se calculent facilement et qui reproduise approximativement les circonstances précédentes. Dans ce but, M. Borel assimile ces décès à des tirages faits dans des urnes où les boules noires par exemple seraient liées par groupes, si bien que le tirage de l'une entraînerait le tirage des autres : le tirage d'une noire ne dépendra plus seulement du nombre des boules de chaque couleur, mais, dans une certaine mesure, du tirage antérieur d'une autre noire.

Supposons, dit M. Borel, qu'une urne contienne 50 boules, dont une noire. Pour représenter le nombre de décès annuel sur une population de 40 millions d'habitants, on compte le nombre de boules noires sorties au cours de 40 millions d'extractions (chaque boule sortie étant remise après qu'on a noté sa couleur). Il y aura en moyenne $\frac{40.000.000}{50} = 800.000$ décès par an.

Supposons maintenant que, sans changer d'urne, on convienne de compter la noire pour 1.000 décès : le

nombre de boules noires extraites représentera le nombre de milliers de décès par an. Pour avoir le même nombre moyen de décès, il suffira de faire 40.000 tirages, ce qui donnera 800 extractions de noires, et par suite 800.000 décès.

Quel sera l'écart qu'on pourra prévoir dans les deux cas ? Dans le deuxième cas, l'écart quadratique moyen du nombre des extractions de boules noires sera : $\sqrt{npq}$, et celui du nombre des décès : $\mu = D\sqrt{npq}$, D désignant le nombre de décès représenté par chaque boule noire. Dans les deux cas :

$$p = \frac{1}{50}, \quad q = \frac{49}{50}.$$

Dans le premier cas :

$$D = 1 \,;\; n = 40.000.000 \,;\; \mu' = \sqrt{40.000.000 \times \frac{49}{50^2}} = \frac{7000}{50}\sqrt{40}$$

Dans le deuxième cas :

$$D = 1.000 \,;\; n = 40.000 \,;\; \mu = 1.000\sqrt{40.000 \times \frac{49}{50^2}}$$

$$= \frac{7.000}{50}\sqrt{40.000}.$$

Le rapport de ces deux nombres est :

$$\frac{\mu}{\mu'} = \frac{\sqrt{40.000}}{\sqrt{40}} = \sqrt{1000} = 31,6,$$

Donc, le nombre moyen de décès étant le même, l'écart quadratique moyen du nombre de décès sera

environ 31 fois plus grand dans le second cas que dans le premier ; le coefficient de dispersion $Q = 31{,}6$; la loi de distribution est hypernormale.

M. Borel étudie aussi l'exemple suivant :

Application aux naissances doubles. — Nous avons vu que la fréquence des naissances masculines était à peu près constante (égale à environ 0,515). En supposant que les sexes des jumeaux soient indépendants, on pourrait chercher les fréquences des diverses combinaisons possibles. Il y en a trois : deux jumeaux peuvent être deux garçons (GG), deux filles (FF), un garçon et une fille (GF). Soient g, f, h, les probabilités respectives (on a évidemment, puisque ce sont tous les cas possibles, $g + f + h = 1$). Soit p la probabilité qu'un nouveau-né pris au hasard soit un garçon.

En appliquant le théorème des probabilités composées, on aurait :

$$g = p^2 ; \; f = q^2 ; \; h = 2pq \text{ en posant } q = 1 - p.$$

Comme nous savons que p est voisin de $\frac{1}{2}$, mais légèrement supérieur, on aurait, en posant :

$$p = \frac{1}{2} +$$

ε étant petit et positif :

$$q = \frac{1}{2} - \varepsilon ; \; g = \left(\frac{1}{2} + \varepsilon\right)^2 ; \; f = \left(\frac{1}{2} - \varepsilon\right)^2 ;$$
$$h = 2\left(\frac{1}{2} + \varepsilon\right)\left(\frac{1}{2} - \varepsilon\right).$$

Ainsi g et f devraient être voisins de $\frac{1}{4}$, le premier plus grand, le second plus petit, et h devrait être légèrement inférieur à $\frac{1}{2}$. On aurait donc :

$$f + g = \frac{1}{2} + 2\varepsilon^2, \qquad h = \frac{1}{2} - 2\varepsilon^2,$$

d'où

(42) $$h > g > f \quad g + f > \frac{1}{2} > h,$$

ces quantités $g + f$ et h étant peu différentes de $\frac{1}{2}$.

Si même on fait le calcul de f, g, h, en prenant $p = 0{,}515$, on trouve les valeurs approchées de g, f, h.

$$g'' = (0{,}515)^2 = 0{,}265\,;\ f'' = (0{,}485)^2 = 0{,}235\,;$$
$$h'' = 2 \times 0{,}515 \times 0{,}485 = 0{,}499.$$

Si l'on se reporte aux résultats donnés par la Statistique générale de la France en 1905, on trouve, pour N couples de jumeaux, des fréquences g', f', h', telles que :

$$\begin{array}{lll} Ng' = 2.933 & \text{couples} & \text{GG} \\ Nf' = 2.797 & \text{—} & \text{FF} \\ Nh' = 3.188 & \text{—} & \text{FG et GF.} \\ N = \text{Total} = 8.918. & & \end{array}$$

D'où l'on tire :

$$g' = \frac{2.933}{8.918} = 0{,}328\,;\ f' = \frac{2.797}{8.918} = 0{,}314\,;\ h' = \frac{3.188}{8.918} = 0{,}358.$$

On a bien, comme pour les inégalités (42) :

$$h' > g' > f'\,;\ g' + f' = 0{,}642 > \frac{1}{2} > 0{,}358 = h'$$

mais on ne peut dire que f', g', sont voisins de $\frac{1}{4}$, et et h', de $\frac{1}{2}$.

Il apparaît donc qu'on ne peut supposer que les deux naissances successives soient deux événements indépendants. Non que le sexe de la première naissance « détermine » nécessairement le sexe de l'autre, mais il peut exister entre les sexes des deux naissances une « corrélation » telle que, dans une proportion définie des cas, ces deux sexes soient nécessairement identiques. En effet, puisque les deux naissances se produisent au même moment, les circonstances qui déterminent le sexe (par exemple, le degré de maturation de l'œuf au moment où il est fécondé) paraissant identiques, il serait naturel de supposer que les sexes doivent être fréquemment les mêmes.

M. Borel représente schématiquement la détermination des sexes par le jeu de hasard suivant.

Dans une urne, un nombre F_1 des B boules sont marquées F, et G_1 boules sont marquées G ; mais une certaine proportion λ d'entre elles ont été marquées deux fois, de sorte que λF_1 boules sont marquées FF, et λG_1 boules sont marquées GG. On convient que, les tirages s'effectuant en remettant chaque boule tirée après avoir noté les lettres inscrites, on fera les tirages par paires, mais en se contentant du premier tirage lorsque celui-ci donne une double lettre. Alors les probabilités g, f, h, seront, d'après le théorème des probabilités totales et composées :

$$
\begin{aligned}
f &= \lambda \frac{F_1}{B} + (1 - \lambda) \frac{F_1}{B} \times \frac{F_1}{B} \\
g &= \lambda \frac{G_1}{B} + (1 - \lambda) \frac{G_1}{B} \times \frac{G_1}{B} \qquad (43) \\
h &= 2 (1 - \lambda) \frac{F_1}{B} \times \frac{G_1}{B} \cdot
\end{aligned}
$$

Posons :

$$\frac{G_1}{B} = p, \frac{F_1}{B} = q;$$

puisque $B = G_1 + F_1$, on a évidemment $p + q = 1$, donc :

$$f = \lambda q + (1 - \lambda) q^2 = q^2 + \lambda pq;$$
$$g = \lambda p + (1 - \lambda) p^2 = p^2 + \lambda pq; \; h = 2 (1 - \lambda) pq.$$

On peut remplacer les deux premières égalités par les égalités équivalentes :

$g - f = p^2 - q^2 = (p - q)(p + q) = p - q = 2p - 1$, (puisque $p + q = 1$) et $g + f = p^2 + q^2 + 2\lambda pq$.

Mais nous avons vu qu'on a nécessairement, dans toutes les hypothèses, $g + f + h = 1$, ou $h = 1 - (g + f)$.

D'où :

$h = 1 - p^2 - q^2 - 2\lambda pq = (p + q)^2 - p^2 - q^2 - 2\lambda pq$
$= 2 (1 - \lambda) pq$.

De sorte que la troisième égalité $h = 2 (1 - \lambda)pq$ est équivalente à la deuxième.

En résumé l'hypothèse n'est acceptable que si

$$g - f = 2p - 1$$

et alors la proportion λ sera donnée par la formule

$$h = 2(1-\lambda)\,pq \quad \text{ou :} \quad 1-\lambda = \frac{h}{2p(1-p)}\,.$$

Remarquons d'ailleurs que $p = \frac{G_1}{B}$ donne la proportion des garçons dans les naissances simples ou multiples.

Nous avons donc d'abord à chercher si la condition : $g - f = 2p - 1$, ou :

$$(44) \qquad p = \frac{1}{2} + \frac{g-f}{2}$$

est vérifiée d'une façon suffisamment approchée quand on substitue à p, g, f les valeurs des fréquences correspondantes observées. Si on prend, comme en 1905 pour la France, $g' = 0{,}328$, $f = 0{,}314$, on aura : $g' - f = 0{,}014$, de sorte qu'on obtiendra pour valeur appprochée de la proportion p des garçons : $\frac{1}{2} + 0{,}007$; quantité qui en effet diffère peu de la proportion des naissances masculines observée. Et la condition (44) sera aussi vérifiée assez bien dans toute statistique où g et f sont peu différentes, quelles que soient leurs valeurs.

Dans ces conditions nous pouvons poursuivre, et calculer λ, par la formule approchée $\lambda = 1 - 2h$, puisque, p étant voisin de $\frac{1}{2}$, $2p(1-p)$ est encore plus voisin de $\frac{1}{2}$, d'où $\lambda = 0{,}28$.

Ainsi tout se passe comme si, 28 fois sur 100, le sexe du second jumeau était déterminé par le sexe du premier.

Remarque. — Il est intéressant d'appliquer la discussion précédente à des nombres statistiques différents de ceux utilisés par M. Borel. Considérons donc les chiffres relatifs à l'Alsace et à la Lorraine pendant la période 1913-1918. On sera surpris de voir combien les résultats relatifs à cette époque troublée diffèrent peu des précédents.

Les nombres Ng', Nf', Nh', N sont ici 726 ; 635 ; 851 ; 2.212 d'où :

$$g' = 0{,}328 \,;\; f' = 0{,}288 \,;\; h' = 0.385.$$

On a bien encore

$$h' > g' > f' \,;\; g' + f' > \frac{1}{2} > h'$$

mais là aussi $g' + f = 0{,}616$ et $h' = 0{,}385$ sont notablement différents de $\frac{1}{2}$.

La valeur $\frac{1}{2} + \frac{g' - f'}{2} = 0{,}52$ reste encore une valeur approchée de p, de sorte qu'on peut calculer λ par la formule $\lambda = 1 - 2h'$; d'où : $\lambda = 0{,}23$, valeur peu différente de la précédente :

Le sexe de la première naissance entraînerait 23 fois (au lieu de 28) sur 100 celui de la seconde naissance.

Schéma des urnes de Borel. — Si nous considé-

rons les équations (43), nous voyons qu'on peut les écrire sous la forme :

$$g = \mu p_1 + \lambda p_2, f = \mu p_3 + \lambda p_4, h = \mu p_5 + \lambda p_6.$$

Demandons-nous alors, avec M. Borel, si l'on ne pourrait représenter sous forme abrégée un grand nombre de tableaux statistiques en déterminant un certain nombre de probabilités . $p_1, p_2, \ldots$, et de coefficients numériques : λ_1, λ_2, ... de telle sorte que les données du tableau résultent de la combinaison des fluctuations autour des probabilités p_1, p_2 ..... des fréquences correspondantes. Si l'on a pu établir un tel système p_1, p_2, ... λ_1, λ_2, ... et vérifier la concordance des résultats statistiques (dans la limite des erreurs prévues par la loi de Laplace) avec les résultats théoriques déduits du tableau, on aura là non pas, peut-être, une explication des résultats expérimentaux passés, mais un moyen de prévoir les résultats futurs.

5^e^ Section

Distinction des cas où les écarts observés peuvent être légitimement attribués aux effets du hasard. — Nous allons rechercher jusqu'à quel point certains écarts constatés dans les séries statistiques sont compatibles avec les résultats auxquels nous sommes parvenus.

Tout d'abord remarquons que si l'on a constaté une fréquence f d'un événement dans un groupe de n épreuves, on peut chiffrer l'intervalle où il est raisonnable de placer la probabilité correspondante.

En effet l'écart quadratique moyen des fréquences dans les séries de n épreuves est $\mu = \sqrt{\frac{pq}{n}}$; et on peut y remplacer approximativement pq par $f\ (1 - f)$. Or la probabilité d'un écart trois fois égal à l'écart quadratique moyen est très petite, moins de $\frac{1}{300}$ d'après la table numérique de la page 290. Ainsi un tel écart ne se présentera en moyenne qu'une fois sur trois cents. On peut le considérer comme pratiquement impossible, de sorte que l'inégalité $|p - f| > 3\sqrt{\frac{f\,(1-f)}{n}}$ est pratiquement impossible, et on pourra affirmer avec une grande certitude que p est compris entre

$$f - 3\sqrt{\frac{f\,(1-f)}{n}} \quad \text{et} \quad f + 3\sqrt{\frac{f\,(1-f)}{n}}\,[1].$$

On peut encore considérer la question à un autre point de vue, et se contenter d'évaluer, pour une valeur

1. Lorsque n n'est pas très grand, il est dangereux de remplacer $p\ (1 - p)$ par $f\ (1 - f)$. Ainsi, pour $n = 1$ et $f = 0$, la condition $|p - f| \leq 3\sqrt{\frac{p\,(1-p)}{n}}$ est vérifiée pour $p = \frac{9}{10}$. Mais on peut remarquer qu'on a toujours $p\ (1 - p) \leq \frac{1}{4}$, et se contenter de la condition $|p - f| < \frac{3}{2\sqrt{n}}$.

convenable de t, la probabilité d'un écart $|p - f| \leq t\sqrt{\frac{f(1-f)}{n}}$. Cette probabilité sera donnée approximativement, si n est assez grand, par la valeur $\Psi(t)$ donnée par la table en face de t (ou obtenue par interpolation).

Par exemple, Buffon, ayant jeté une pièce de monnaie 4.040 fois, a obtenu face 2.048 fois ; dans ce cas $f = \frac{2.048}{4.040}$ et $n = 4.040$. (Si la fréquence avait été $\frac{1}{2}$, il aurait obtenu face 2.020 fois.) La probabilité d'obtenir une telle fréquence à partir de la probabilité $p = \frac{1}{2}$ est $\Psi(t)$, t étant donné par

$$\left| \frac{1}{2} - \frac{2.048}{4.040} \right| = t\sqrt{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{4.040}} = \frac{t}{2\sqrt{4.040}}.$$

$$t = \frac{56}{\sqrt{4.040}} = 0{,}88.$$

On trouve $\Psi(t) = 0{,}622$, ; autrement dit, si l'on fait un grand nombre N de séries de 4.040 jets d'une pièce de monnaie régulière, on peut s'attendre à trouver environ 378 séries sur 1.000 où l'on aurait obtenu un nombre de fois face au moins aussi différent de 2020 que dans le cas de Buffon.

Mais il faut bien remarquer que le raisonnement précédent suppose essentiellement l'existence d'une probabilité constante dans chaque épreuve, et d'une probabi-

lité à laquelle s'appliquent les lois des probabilités totales et composées. Si, par exemple, la pièce de monnaie était en plomb, si elle était jetée avec force dans une direction convenable sur un sol dur, il pourrait arriver qu'elle se déformât de plus en plus, et que la probabilité p s'éloignât de plus en plus de $\frac{1}{2}$.

En d'autres termes la formule (p. 285) précise quel est l'écart entre la fréquence cherchée et la probabilité admise en supposant le jeu naturel des lois du hasard. Si l'écart est plus grand, il n'est pas impossible, mais il apparaît extrêment improbable, et nous devons alors nous demander soit si nous ne nous sommes pas trompés dans nos calculs, soit si les hypothèses que nous avons faites sur l'existence et la constance de la probabilité étaient exactes.

Exemple. — Nous avons étudié à la page 263 la fréquence annuelle des naissances de garçons dans $N = 26$ années consécutives, pour une moyenne annuelle de $n = 734.449$ naissances. Remarquons d'abord que la fréquence moyenne constatée $f = 0{,}515$ sera égale à la probabilité supposée constante à une erreur près inférieure à $3\sqrt{\frac{f(1-f)}{Nn}}$, car ici f est non seulement la moyenne des 26 fréquences, mais aussi la fréquence dans l'ensemble des $Nn = 26 \times 734.449$ naissances. Donc :

$$|p - 0{,}515| < 3\sqrt{\frac{0{,}515 \times (1 - 0{,}515)}{26 \times 734.449}} < 0{,}0004 ;$$

$0{,}5146 < p < 0{,}5154$.

Telle serait la précision avec laquelle p serait connu en admettant une probabilité constante.

Vérification expérimentale. — Charlier a fait 10.000 tirages de cartes extraites, notées et replacées dans un paquet de cartes contenant autant de cartes noires que de cartes rouges. Il les a rangées par groupes de 10 tirages, et a noté chaque fois le nombre $m = 10\,f$ de cartes noires extraites ; le nombre de groupes ayant donné la valeur m étant z, on a :

$10\,f = m$	0	1	2	3	4	5	6	7	8	9	10	TOTAL.
z	3	10	43	116	221	247	202	115	34	9	0	1.000

on en tire, pour la valeur moyenne de m : 4,933 ; d'où : $f_0 = 0{,}4933$.

La probabilité p serait donc comprise entre

$$f_0 \pm \theta \sqrt{\frac{f_0\,(1 - f_0)}{10.000}}$$
$$= 0{,}4933 \pm \theta \times 0{,}005$$
$$0{,}4933 - \theta \times 0{,}005 < p < 0{,}4933 + \theta \times 0{,}005.$$

Si on prend $\theta = 3$, puisqu'il y a une probabilité très petite 0,003 pour que $\theta > 3$, on aura

$$0{,}4783 < p < 0{,}5083.$$

Il n'y a donc pas contradiction à prendre ici $p = \frac{1}{2}$, la valeur obtenue en considérant le tirage de

chaque carte comme également possible. — On trouve aussi pour l'écart quadratique moyen

$$\mu'_0 = \sqrt{\frac{\Sigma\,(f_1 - f_0)^2}{N}} = \sqrt{\frac{24,15}{1.000}} = 0,04915.$$

(Pour la quantité $\mu'_0 = \sqrt{\frac{\Sigma(f_1 - f_0)^2}{N - 1}}$, la quantité $\mu'_0 = \mu'_0 \sqrt{\frac{N}{N - 1}}$ ne diffère de μ'_0 que par la décimale suivante).

Le coefficient de dispersion est 0,983, qui est en effet voisin de l'unité.

TABLE DES VALEURS DE LA FONCTION DE LAPLACE

Rappelons que nous avons appelé, page 222, $\Psi(t)$ la probabilité pour qu'un écart l soit en valeur absolue supérieur à t fois l'écart quadratique moyen μ de l (rigoureusement, $\Psi(t)$ n'est en réalité que la limite de cette probabilité pour un nombre croissant d'épreuves).

Pour la pratique, la table ci-dessous sera généralement suffisante[1].

t	0	1	2	3	4	5	6	7	8	9
0,	0,000	0,080	0,159	0,236	0,311	0,383	0,451	0,516	0,576	0,632
1,	683	729	770	806	838	866	890	911	928	943
2,	954	964	972	978	984	988	991	993	995	996
3,	997	998	999	999	999					

Exemples $\Psi(0,2) = 0,159$, $\Psi(2,0) = 0,954$.

1. On peut définir analytiquement $\Psi(t)$ par la formule

Les nombres de la table ci-dessus sont exacts à une demi-unité près du troisième ordre, c'est-à-dire que par exemple la vraie valeur de Ψ (3, 4) est comprise entre 0,9985 et 0,9995 (celles de Ψ (3,5), Ψ (3,6),..... sont supérieures à 0,9995, et naturellement inférieures à l'unité).

Valeurs intermédiaires. — La méthode des parties proportionnelles permettra de déterminer les valeurs de Ψ (t) non portées dans la table à moins d'un centième près : autrement dit on pourra ainsi obtenir à moins d'une unité près le pourcentage des écarts $|l|$ supérieurs à $t\mu$, pour toute valeur de t.

Exemple. — Soit à évaluer Ψ (1,165). Cette valeur sera comprise entre Ψ (1,1) $=$ 0,729 et Ψ (1,2) $=$ 0,770 ; désignons-la par $0{,}729 + \varepsilon$. L'accroissement ε de Ψ quand t s'accroît de 0,065 à partir de 1,1 devient $0{,}770 - 0{,}729 = 0{,}041$ quand t s'accroît de 0,1 en passant de 1,1 à 1,2. Si on suppose approximativement ces accroissements proportionnels, on a $\frac{\varepsilon}{0.065} = \frac{0{,}041}{0{,}1}$, d'où $\varepsilon = \frac{0{,}065 \times 0{,}041}{0.1} = 0{,}02665$. D'où

$\Psi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{\frac{-u^2}{2}} du$. De sorte que le calcul de Ψ (t) se ramène immédiatement à celui de l'intégrale de Laplace $\int_x^{\infty} e^{-u^2}\, du$ dont la première table a été dressée, à l'instigation de Laplace, par un de nos prédécesseurs à l'Université de Strasbourg, Ch. Kramp, dans son *Analyse des réfractions astronomiques*, Strasbourg, an VII.

approximativement $\Psi(1,165) = 0,729 + 0,02665 = 0,75565$. On peut garantir que la méthode fournit un résultat exact à $\frac{1}{100}$ près pour toute valeur de t. Mais l'exactitude sera généralement plus grande. Comme elle ne peut aller au delà de celle de la table qui est son point de départ, on devra donc laisser de côté les décimales au delà de la troisième (en augmentant celle-ci d'une unité si la quatrième est supérieure à 5).

On écrira donc finalement $\Psi(1,165) = 0,756$, sans garantir le troisième chiffre qui sera cependant souvent exact ou peu éloigné du chiffre exact.

INDICATIONS BIBLIOGRAPHIQUES

Nous signalerons seulement les ouvrages suivants, en commençant par les plus accessibles.

I. — *Essai philosophique sur les probabilités*, par Pierre-Simon, marquis de Laplace, 1814, réédition en deux petits volumes, chez Gauthier-Villars, 1921.

Exposition de la théorie des chances et des probabilités, par A. Cournot, Paris, 1843.

Le hasard, par Emile Borel, chez Alcan, 5e édition, 1922.

Ces trois ouvrages donnent un aperçu général et permettent en particulier d'apprécier la variété des applications du Calcul des Probabilités.

II. — Le lecteur qui s'intéresse à la théorie des jeux en trouvera un bon résumé dans :

Leçons élémentaires sur le Calcul des Probabilités, par de Montessus, Gauthier-Villars, 1908.

III. — *Eléments de la théorie des probabilités*, par E. Borel, chez Hermann, 1910.

Cet ouvrage peut être considéré comme une introduction aux grands traités qui suivent.

Théorie analytique des probabilités, par Laplace, 1812. Réimprimé dans la collection des œuvres de Laplace.

Calcul des probabilités, par Joseph Bertrand, 1889, chez Gauthier-Villars.

Calcul des probabilités, par Henri Poincaré, 1896, chez Naud (2^{e} édition 1912).

Ces trois ouvrages supposent une éducation mathématique assez avancée. Nous les mentionnons en raison des vues nouvelles qu'ils ont successivement introduites dans la science du hasard.

IV. — Parmi les ouvrages étrangers nous recommandons particulièrement :

Le traité en italien : *Calcolo della Probabilità*, par Guido Castelnuovo, chez Albrighi, 1919, où l'auteur a su avec une clarté et une précision remarquables incorporer à la théorie classique de Laplace les découvertes les plus modernes.

Une intéressante étude philosophique : *The logic of chance*, an essay on the foundations and province of the theory of probability, par John Venn ; 1888 (3^{e} édition très remaniée).

Et une histoire très étendue, mais qui date un peu, du développement du Calcul des Probabilités : *A History of the mathematical Theory of probability*, from the time of Pascal to that of Laplace, par I. Todhunter ; Macmillan, 1865.

V. — Enfin on trouvera un certain nombre de tables numériques utiles dans l'ouvrage suivant :

Houël, *Recueil de formules et de tables numériques*, Paris, Gauthier-Villars, 3^{e} édition, 1901.

TABLE DES MATIERES

(Les nombres ci-dessous désignent les numéros des pages.)

ÉVREUX, IMPRIMERIE CH. HÉRISSEY.